University of Alberta

Analytical Technology and Methods for Proteome Analysis Using Chromatographic Media for Protein Capture and Digestion

by

David Robert Craft

©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Department of Chemistry

Edmonton Alberta

Spring 2004

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 0-612-96254-7
Our file Notre référence
ISBN: 0-612-96254-7

The author has granted a non-exclusive license allowing the Library and Archives Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou aturement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Abstract

Mass spectrometry (MS) has been the method of choice for protein identification and characterization for the past several years. Currently, numerous research efforts are focused on proteome analysis. The proteome is defined as the set of proteins being expressed in a cell, tissue, or organism under defined conditions at a specific time, and may consist of several thousand proteins. Therefore, the demand for rapid analysis has recently increased. Mass spectrometers are capable of generating data quickly and thus have a great potential for high-throughput analysis.

In this work a new technology has been developed for protein identification using mass spectrometry that incorporates sample cleanup, preconcentration, and protein digestion in a single-stage system. The procedure involves the adsorption of a protein, or protein mixture, from solution onto a hydrophobic media that is contained within a microcolumn. The protein is digested while still bound to the hydrophobic support. The method is demonstrated with standard protein samples at concentrations down to 10 nM. Real world samples, as well as hydrophobic proteins, are also successfully digested on the hydrophobic media and detected by mass spectrometry. Peptide fragments, generated from digestion of standard protein samples bound to various types of micro-bead surfaces, were examined to determine the effects of surface chemistry and surface morphology on the digestion process. It is demonstrated that digesting proteins on various types of hydrophobic surfaces produce peptide mass fingerprints with only minor differences.

Finally, this technology is adapted to directly couple the surface digestion procedure with mass spectrometric detection. The entire system is fully automated and

optimization for high throughput analysis. Proteins extracted from *E. coli* cells were used to demonstrate the power of this novel technology. The methods and system described in this thesis provide reliable protein identification capable of high-throughput analysis. This technology offers a potential solution for high throughput proteome analysis.

for my wife, Janelle

Acknowledgements

First of all, I would like to thank my supervisor, Dr. Liang Li, for his advice, support, and guidance over the course of my research at the University of Alberta. His ideas, encouragements, suggestions, and enthusiasm were always appreciated, and I feel that I have learned a significant amount from him.

I would also like to thank the many members of Dr. Li's research group that I have worked with throughout my research. In particular, I would like to thank Dr. Alan Doucette for his contributions. I especially appreciate Dr. Doucette taking the time to proof read my thesis and providing helpful comments throughout. I would also like to thank Chris McDonald for proof reading my thesis.

I would like to thank the Department of Chemistry and the personnel in the machine shop and the electronic shop of the department.

I would like to thank my parents, family, and friends for there tremendous support throughout my degree.

Finally, I would like to thank my wife, Janelle, for providing me with her support, attention, acceptance, and most of all, patience, over the course of my studies at the University of Alberta. Janelle made several personal sacrifices in order for me to obtain my degree. I recognize and appreciate these sacrifices. Thank you Janelle.

Table of Contents

Chapter 1 Introduction: Mass Sepctrometry Methods for Protein Identification in Proteomics 1

1.1 Advances in Proteomics 1

1.2 Ionization Methods 4

1.2.1 MALDI 4

1.2.1 Electrospray 7

1.3 Mass Spectrometers 11

1.3.1 MALDI Time-of-Flight Mass Analyzer 11

1.3.1.1 Time-Lag Focusing 13

1.3.1.2 Reflectron TOF Instruments 15

1.3.2 Quadrupole Mass Analyzer 15

1.3.3 Quadrupole IonTrap 17

1.4 Protein Identification 19

1.4.1 Peptide Mass Fingerprinting 19

1.4.2 MS Fragmentation of Peptides 20

1.4.3 Database Searching 23

1.5 Sample Preparation and Analysis by Mass Spectrometry 24

1.5.1 Difficulties in Sample Preparation Prior to Mass Spectrometry 24

1.5.1.1 Sample Contamination 24
1.5.1.2 Protein Digestion 25
1.5.2 Sample Preparation Automation 26
1.6 Challenges Associated with Proteome Analysis 27
1.7 Literature Cited 28
Chapter 2 Microcolumn Capture and Digestion of Proteins Combined with Mass Spectrometry for Protein Identification 33
2.1 Introduction 33
2.2 Experimental 36
2.2.1 Materials and Reagents 36
2.2.2 *E. coli* Protein Extraction. 37
2.2.3 Anion Exchange Chromatography. 37
2.2.4 Protein Digestion in Microcolumns 38
2.2.5 MALDI Sample Preparation 39
2.2.6 Mass Spectrometry. 39
2.2.7 Safety Considerations. 40
2.3 Results and Discussion 41
2.3.1 Optimization of Microcolumn Size. 41
2.3.2 In-Column Digests of Standard Proteins. 42

2.3.3 Samples Contaminated with NaCl, Urea, and SDS. 44
2.3.4 Protein Digestion at Shorter Reaction Times. 46
2.3.5 Concentration Limit for In-Column Digestion. 46
2.3.6 Anion Exchange Fractionation of *E. coli*. 49
2.3.7 Analysis of Human Plasma. 54
2.3.8 Analysis of Human Tears 59

2.4 Conclusions 61

2.5 Literature Cited 62

Chapter 3 Mass Spectrometric Study of the Effects of Hydrophobic Surface Chemistry and Morphology on the Digestion of Surface-Bound Proteins 64

3.1 Introduction 64

3.2 Experimental 68
3.2.1 Materials and Reagents 68
3.2.2 Digestion of Cytochrome c Adsorbed on Micro-Beads 69
3.2.3 Solution Digestion of Protein 69
3.2.4 Protein Digestion in Microcolumns 70
3.2.5 HPLC Separations and Fractionation 71
3.2.6 Mass Spectrometry 71

3.3 Results and Discussion 72
3.3.1 HPLC Separation with UV detection of Peptide Fragments 72

3.3.2 MS Detection of Solution and Surface-Bound Protein Digests 77
3.3.3 In-Column Digestion of Proteins on Various Supports................................ 84
3.3.4 Effects of Bead Construction and Pore Size on Digestion 89
3.3.5 Enzymatic and Chemical Cleavage of Proteins Adsorbed on Microbeads .. 90
3.3.6 Digestion of Hydrophobic Protein.. 94

3.4 Conclusions.. 95

3.5 Literature Cited 98

Chapter 4 An Automated Approach to Surface Digestion ... 101

4.1 Introduction.. 101

4.2 Experimental 104
4.2.1 Materials and Reagents... 104
4.2.2 Microcolumn Design 104
4.2.3 Sample Manipulation.. 105
4.2.3.1 Loading and Washing 105
4.2.3.2 On-line Protein Digestion 105
4.2.3.3 Elution of Peptide Fragments.. 107
4.2.4 Configuration of On-line Digestion System 107
4.2.5 Instrumentation 111
4.2.6 SEQUEST Analysis.. 111

4.3 Results and Discussion 112

4.3.1 Optimization of System 112

4.3.2 On-line Digestion of Proteins 115

4.3.3 Effect of Temperature on Digestion 118

4.4.4 Re-use of Digestion Column 118

4.4 Conclusions 123

4.5 Literature Cited 124

Chapter 5 Automated Online Digestion of *Escherichia coli* Proteins 127

5.1 Introduction 127

5.2 Experimental 131

5.2.1 Materials and Reagents 131

5.2.2 Growth of *E. coli* 131

5.2.3 Cell Lysate 132

5.2.4 Separations 133

5.2.4.1 Anion Exchange Chromatography 133

5.2.4.2 Reversed-Phase Chromatography 133

5.2.5 Sample Manipulation on Microcolumn 134

5.2.6 Solution Digestion 134

5.2.7 Instrumentation 134

5.3 Results and Discussion 135

5.3.1 Pre-Fractionation Prior to Digestion 135

5.3.2 Digestion of Anion Exchange Fractions 137

5.3.3 Analysis 140

5.3.4 Method Comparison 143

5.4 Conclusions 146

5.5 Literature Cited 147

Chapter 6 150

Conclusions and Future Work 150

Appendix A 162

List of Tables

Table 2.1 Summary of the MASCOT search results for peptides generated from a digest of an *E. coli* fraction, obtained from an anion exchange separation of the total cell extract. 53

Table 2.2 Summary of the polypeptides identified using MS/MS data from a diluted human plasma fraction, as obtained from the SEQUEST Search Algorithm. 57

Table 2.3. List of the four most abundant proteins in tears. 60

Table 3.1 Tryptic peptides detected from the digests of cytochrome c bound to various surfaces. 87

Table 3.2 Tryptic peptides detected from the digests of BSA bound to various surfaces. 88

Table 3.3 Summary of different enzymes and chemicals used to degradation of proteins to peptides. 91

Table 4.1 Summary of switching valve positions for each cycle of the system. ..110

Table 4.2 Summary of gradients used for each HPLC pump. 110

Table 5.1 Bradford protein assay results from anion exchange fractions. 139

List of Figures

Figure 1.1 Schematic representation of matrix-assisted laser desorption ionization (MALDI). Matrix molecules (blank circles) are mixed with the analyte (black ovals) on a stainless steel MALDI target. A laser pulse is applied to the mixture to induce ionization. 5

Figure 1.2 Schematic representation of the electrospray process. First charged droplets are produced at the capillary tip. Solvent evaporation causes shrinkage of the droplets resulting in repeated droplet disintegration until gas phase ions are produced. 8

Figure 1.3 ESI mass spectrum of lysozyme. Several multiple charged ions in the spectrum are observed, creating a charged envelope for lysozyme. The charge state of each ion is displayed above its m/z. 10

Figure 1.4 Schematic of time-lag-focusing. (A) Ions are desorbed from MALDI target with different initial kinetic energy. (B) Ions are expanded in the field free source region where high energy ions move further away from the repeller. (C) A pulse voltage (E) is applied to compensate for the differenced in initial kinetic energy. . (D) Ions reach the detector simultaneously. P1 and P2 are the DC voltages applied on the repeller and extractions grids. 14

Figure 1.5 Schematic diagram of a reflectron time-of-flight mass spectrometer. V is the voltage applied to each plate in the ion mirror. $V_1 > V_2 > V_3 > V_n$. ..16

Figure 1.6 Schematic of a quadrupole ion trap mass spectrometer. The ion trap collects ions and then can be used to select an ion of interest and discard the rest. The ion of interest can then be fragmented by introducing a collision gas (He). The fragmented ions are then ejected out of the trap for detection. 18

Figure 1.7 Nomenclature system for fragment ions generated by collisional induced fragmentation of peptides. The ions labeled a, b, and c refer to charge retention on the N terminus, whereas x, y, and z ions retain the charge on the C terminus. 21

Figure 2.1 MALDI spectra showing the products of the digestion of 100 nM cytochrome c incolumn that was packed with R2 beads for (A) a clean solution, 28 peaks were detected, giving 91% coverage, (B) a solution containing 2M NaCl, 27 peaks were detected, giving 100% coverage, (C) a solution containing 2M urea, 26 peaks were detected, giving 97% coverage, and (D) with 0.01% SDS in solution, 25 peaks were detected, giving 97% coverage. C = peptide fragments of cytochrome c, T = peptide fragments from trypsin autolysis. 43

Figure 2.2 MALDI spectra showing the products of digestion of 100 nM lysozyme in-column packed with R2 beads using (A) 30 min digestion time and (B) 5 min digestion time. L = peptide fragments of lysozyme, T = peptide fragments from trypsin autolysis. Insert displays the MALDI spectrum from 2500 Da to 4500 Da. 47

Figure 2.3 MALDI spectrum obtained from the digestion of 25 nM BSA in-column. B = peptide fragments of BSA, T = peptide fragments from trypsin autolysis. 48

Figure 2.4 Total MS scans from MALDI analysis of the digestion products of an *E. coli* fraction from an ion exchange separation on a hybrid quadrupole-time-of-flight system. (A) is for the *E. coli* sample that was digested in-column and (B) is the same *E. coli* fraction digested using a traditional solution phase protocol, with ZipTip cleanup. CAT = peptide fragments identified by MS/MS fragmentation as *E. coli* chloramphenycol acetyltransferase, K = keratin fragments. 51

Figure 2.5 Base peak chromatogram obtained from the analysis of a diluted plasma fraction following digestion by LC/ESI. Two representative MS/MS spectra are also shown, along with the tentative sequence of the fragmented peptide, and identity of the parent protein. The plasma sample was digested with trypsin using the in-column digestion procedure. 56

Figure 2.6 Base peak chromatogram obtained from the analysis of a diluted tears from human following in-column digestion by LC/ESI. 60

Figure 3.1 UV chromatograms for the HPLC separations of peptide fragments from the digestion of cytochrome c: (A) in solution, (B) on C_{18} silica beads, and (C) on polymeric R2 beads. 74

Figure 3.2 Tryptic peptides detected from the digest of cytochrome *c* in solution by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a gray bar, while those exclusive to LC/off-line MALDI are indicated with a checkered bar. Fragments commonly observed in each MALDI analysis method is indicated with a black bar. 78

Figure 3.3 Tryptic peptides detected from the digest of cytochrome c bound to the C_{18} surface by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI is indicated with a checkered bar. Fragments commonly observed in each MALDI analysis method are indicated with a black bar. 79

Figure 3.4 Tryptic peptides detected from the digest of cytochrome c bound to the R2 surface by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI are indicated with a checkered. Fragments commonly observed in each MALDI analysis method is indicated with a black bar. 80

Figure 3.5 MALDI spectra showing the in-column digestion of 100-nM cytochrome c, where the column was packed with (A) Vydac C_{18}, (B) C_8. (C) C_4, and (D) Poros R2 beads. In the spectra, the label "C" refers to peptide fragments of cytochrome c, and "T" refers to peptide fragments resulting from trypsin autolysis. 85

Figure 3.6 MALDI spectra showing the chymotrypsin digestion of 500-nM BSA on Poros R2 beads packed in a column. In the spectra, the label "*" refers to peptide fragments of BSA. Insert displays MALDI spectra from 2000 Da to 5000 Da. 91

Figure 3.7 MALDI spectra showing the pepsin digestion of 500-nM cytochrome c on Poros R2 beads packed in a column. In the spectra, the label "*" refers to peptide fragments of cytochrome c. 93

Figure 3.8 MALDI spectra showing the CNBr cleavage of 500-nM β-Casein on Poros R2 beads packed in a column. In the spectra, the label "*" refers to peptide fragments of β-Casein. 93

Figure 3.9 MALDI spectra showing the products of (A) in-column digestion of 500 nM bacteriorhodopsin by a combined CNBr/trypsin and (B) in-solution digestion of 2 μg/μL bacteriorhodopsin with CNBr/trypsin. Br = peptide fragments of bacteriorhodopsin, T = peptide fragments from trypsin autolysis. 96

Figure 4.1 Illustrated representation of micro column used for digestion. The column contains 40 μg of Poros R2 beads. The heating wire was heated to 37°C. 106

Figure 4.2 Schematic representation of the on-line sample manipulation workstation. If all three switching valves are diverting flow in the black position then HPLC pump 1 is loading and digesting micro column A. HPLC pump 2 is eluting peptides from micro column B into separation column B and into the LCQ^{DECA} for detection. The third HPLC pump is washing separation column A. Once the cycle is completed all three switching valves are changed to the white position. 109

Figure 4.3 Base peak chromatogram of 100 nM cytochrome c digested on-line at 37^{o}C. The peaks labeled "C" represent cytochrome c peptides as the most intense mass spectral peak. Peaks labeled "T" represents trypsin autolysis peptides as the most abundant mass spectral peak.. 116

Figure 4.4 Fragmentation pattern of cytochrome c peptide 1352 Da. Ions are labeled with the conventional nomenclature system (Figure 1.7). 117

Figure 4.5 SEQUEST search results for cytochrome c peptide with m/z 1352. 119

Figure 4.6 Base peak chromatogram of 100 nM cytochrome c digested on-line at room temperature. 14 peptides unique were correlated to cytochrome c. 120

Figure 4.7 Base peak chromatogram of 400 nM myoglobin digested on-line. A total of 12 unique myoglobin peptides were identified. 122

Figure 4.8 Fragmentation of peptide 3405 Da from the myoglobin digest. This peptide eluted from the C_8 separation column at 72 min. 122

Figure 5.1. SDS PAGE image of proteins extracted from *E. coli.* with 100 mM tris/HCl buffered at pH 7.8. 136

Figure 5.2 UV chromatogram of *E. coli* lysate separated on a Mono Q HR 5/5 column. Fractions 8 – 50 were subsequently digested for protein identification. 138

Figure 5.3 Base peak chromatograms from fractions 9, 22, 29, and 44. A total of 4, 4, 9, 5 proteins were identified in each respective fraction. 141

Figure 5.4 Typical SEQUEST search results for anion exchange fraction digestion on beads in the automated system. Section A lists every MS/MS cycle performed. Section B summarizes the data from section A. 145

Figure 5.5 Base peak chromatogram of fraction 29 minutes, from anion exchange separation. Digestion was performed in solution and concentrated with ZipTip prior to injection. A total of 9 proteins were identified when searched against the ecoli.fasta database. 145

Figure 6.1 Flow chart representing a strategy for protein identification of complex samples. 155

Figure 6.2 With the addition of two switching valves and one pump the ion exchange column can be directly coupled to the automated digestion system. ...157

Figure 6.3 Direct MALDI spectrum of proteins adsorbed to Poros R2 beads from human cells. 160

Figure 6.4 Direct MALDI spectrum of peptides produced from digesting proteins extracted from human cells on Poros R2 beads..............................160

List of Abbreviations

2D	two-dimensional
Å	angstrom
ATCC	American Type Culture Collection
BSA	bovin serum albumin
C	Celsius
CID	collisional induced dissociation
cTnC	cardiac Troponin C
Da	Daltons
DTT	dithiothreitol
E. coli	*Escherichia coli*
ESI	electrospray ionization
EST	expressed sequence tag
FWHM	full width at half maximum
HCCA	α-cyano-4-hydroxycinnamic acid
HPLC	high pressure liquid chromatography
HSA	human serum albumin
i.d.	internal diameter
IMAC	immobilized metal ion affinity chromatography
LC	liquid chromatography
M	Molar
MALDI	matrix-assisted laser desorption/ionization
μg	microgram

mg	milligram
min	minute
µL	microlitre
mL	millilitre
mM	millimolar
MS	mass spectrometry
MS/MS	tandem-mass spectrometry
MudPIT	Multi-Dimensional Protein Identification Technology
m/z	mass to charge ratio
ng	nanogram
nM	nanomolar
o.d.	outer diameter
pmol	picomolar
PAGE	polyacrylamide gel electrophoresis
pI	isoelectric point
PTM	post-translational modification
PVDF	polyvinylidene fluoride
Q	quadrupole
RF	radio frequency
SDS	sodium dodecyl sulfate
SELDI	surface-enhanced laser desorption/ionization
TFA	trifluoroacetic acid
TLF	time-lag focusing

TOF	time-of-flight
UV	ultraviolet
v/v	volume to volume ratio
w/v	weight to volume ratio

Chapter 1

Introduction:

Mass Spectrometry Methods for Protein Identification in Proteomics

1.1 Advances in Proteomics

Proteomics is a newly emerging science, which focuses on characterizing the proteome, *i.e.* the set of proteins being expressed in a cell, tissue, or organism under defined conditions at a specific time [1]. Proteomics research also encompasses the analysis of protein activities, their modifications, subcellular localization, and the interactions of protein complexes. Proteins are of particular interest because they are involved in almost all biological activities. Thus, identification and characterization of proteins can lead to further understanding of a biological system.

The molecular characterization of proteins by amino acid sequencing first began in 1945 [2]. In this historical work, Frederick Sanger, used a combination of enzymes and acid hydrolysis to cut the protein insulin into many small pieces (*i.e.*, peptides). Using what is now know as the Sanger method for protein sequencing, the entire amino acid sequence of both the alpha and beta chains of the insulin protein could be determined. And roughly 10 years after his initial investigations on the N-terminal amino acids of insulin, Sanger finally reported the full sequence of insulin [3].

At roughly the same point in time, Linus Pauling was studying another important protein, hemoglobin. He became aware of a rare, but debilitating blood disease known as sickle cell anemia and hypothesised that hemoglobin was involved in the disease. Pauling attempted to apply Sanger's methods to sequence the hemoglobin protein;

however, due to hemoglobin being significantly larger than insulin, he was unsuccessful in the attempt. He was, however, able to demonstrate that the hemoglobin molecule was different in normal *vs.* sickle cell patients, as described in his groundbreaking paper entitled "Sickle Cell Anemia: A molecular disease" [4]. In 1956, Vernon Ingram used an ingenious method to simplify the process of isolating the molecular difference between normal and sickle cell hemoglobin [5]. Ingram successfully digested hemoglobin with enzymes and separated the peptide fragments using thin layer chromatography. When comparing normal *vs.* sickle cell hemoglobin he noticed that while most of the peptides were in the same position, only 1 was different between the two separations. He proceeded to sequence this unique peptide from each of the two sources using Sanger's method and discovered that they were different. In particular, a single Glycine residue was substituted for Valine in the sickle cell hemoglobin. He termed this technique of protein digestion and peptide display for probing differences on protein structure "peptide fingerprinting", and thus a new tool in protein characterization was born.

One of the most significant breakthroughs in protein characterization came with O'Farell's work, which made a significant improvement in separation techniques in 1973 [6]. Two-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (2D-SDS-PAGE) successfully resolved up to 1000 proteins at a given time. By this point, Edman degradation (a technique similar to Sanger's sequencing method) had matured and became the premier method for sequencing proteins [7]. With this technology, proteins in the high abundant gel bands were successfully sequenced [8]. However, Edman degradation had neither the sensitivity, nor the speed to fully characterize the separated protein set.

The potential of speed and sensitivity did not come until mass spectrometry (MS) was first introduced for protein and peptide detection [9]. The breakthrough came when matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI) by Tanaka and Fenn respectively [10-12] were adapted for the ionization of biomolecules. These soft ionization processes allowed large biomolecules to enter the gas phase with minimal fragmentation, thus making it possible to analyze them by mass spectrometry. Before these techniques, the ionization methods employed for mass spectrometry, such as electron impact, were simply too hard and fragmented the large molecules during the ionization process. The ability to detect low amounts of proteins and peptides using MALDI and ESI was an enormous step in proteomics and was recognized by the Nobel Prize in Chemistry 2002.

Since the advent of MS for protein analysis, numerous advances have been made. These include new developments in computer programs for spectral interpretation as well as linking this information to protein or genome sequence databases. This introduction will focus on improvements on the previously described methods and the new technologies inspired by them. Several mass spectrometers (*i.e.*, time-of-flight, quadrupole, and ion traps) have facilitated the process for protein identification. These instruments allowed the growth of peptide mass mapping and enabled tandem mass spectrometric analysis of peptide ions for generating sequence information. However, the field of mass spectrometry for protein characterization is still in its developmental stage, as numerous challenges still need to be met. Background materials as well as some of the current problems with MS methodologies for proteomic analysis will be discussed below.

1.2 Ionization Methods

1.2.1 MALDI

Matrix-assisted laser desorption ionization (MALDI), a soft (*i.e.*, low energy) ionization method, was introduced in 1987 by two independent groups. In order to generate gas phase ions, a large excess of matrix material was co-precipitated with the analyte. Tanaka *et al.* [11] utilized fine cobalt powder suspended in glycerol as a matrix. With this technique, he was able to analyze high molecular weight proteins (up to 100 kDa). The technique simultaneously discovered by Karas and Hillenkamp [13,14] is the method of choice today. This method was first discovered through work involving the ionization of Tryptophan and Alanine, using laser desorption to generate gas phase ions. In their preliminary experiments, the wavelength of the laser was chosen to desorb only Tryptophan; however, a strong signal for Alanine was also observed. The desorption of Tryptophan carried the Alanine along with it [15]. These observations led to the development of small UV absorbing molecules for desorbing proteins by Karas and Hillenkamp [16,17].

The current technique employs the co-crystallization of matrix (low molecular weight, UV absorbing, proton donating, organic crystalline compound) and analyte on a target plate in a ratio of approximately 500:1 [18]. Figure 1.1 shows a schematic representation of the MALDI process. The figure displays analyte and matrix mixed together on a MALDI target. The target is typically stainless steel; however, several research efforts are directed toward employing alternative materials such as Teflon or other non-metallic materials. Irradiation of these crystals by short pulses of high intensity laser UV light initiates desorption and ionization. The chromophore of the matrix

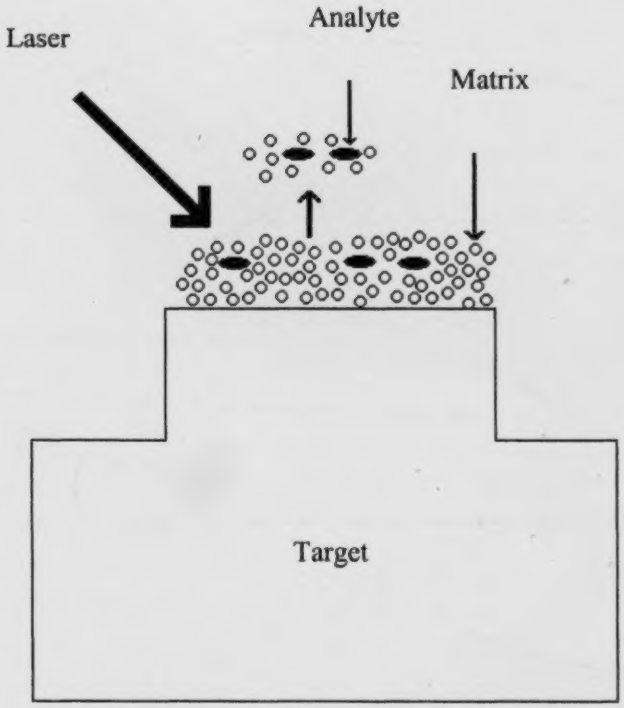

Laser
Analyte
Matrix
Target

Figure 1.1 Schematic representation of matrix-assisted laser desorption ionization (MALDI). Matrix molecules (blank circles) are mixed with the analyte (black ovals) on a stainless steel MALDI target. A laser pulse is applied to the mixture to induce ionization.

molecules absorbs the energy of the laser and expands rapidly into the gas phase, carrying along the analyte. The MALDI ionization mechanism has been debated since this method was introduced [19-23]. However, it is generally accepted that ionization occurs *via* gas phase proton-transfer reactions between excited matrix and analyte molecules [24]. The signal intensities depend on the co-crystallization of the peptides with matrix, their probability of capturing and/or retaining a proton during the desorption process, and a number of other factors including suppression effects in peptide mixtures [25]. For example, peptides with a C-terminal arginine generally give higher signal intensity than peptides with a C-terminal lysine [26]. For these reasons, it is difficult to relate peptide peak area with the absolute quantity of sample present unless an internal standard is used.

MALDI-MS produces predominantly singly-protonated peptide ions ($[M+H]^+$). Proteins, however, can produce several other ionic species, such as $[M+2H]^{2+}$, $[M+3H]^{3+}$ and $[M+M+H]^+$, depending on their size and the type of matrix used. Proteins and peptides are also susceptible to adduct formation with metal ions. This occurs when sodium, potassium or other metal ions present in the sample replaces the hydrogen, *i.e.*, $[M+Na]^+$, $[M+2Na]^{2+}$, $[M+K]^+$, and $[M+Na+K]^{2+}$. The formation of salt adduct ions reduces the protonated analyte signal intensity and produces many unwanted peaks in the mass spectrum which can complicate the spectral interpretation. Therefore, it is crucial to eliminate the salt from the matrix and analyte on the MALDI target. The two-layer method, the crystallization process tends to exclude the inorganic salts, as it precipitates on top of the matrix-analyte crystal, thus facilitating their removal by on-probe washing with water.

The mass range below approximately 500 daltons (Da) in a MALDI spectrum exhibits the typical low-mass background of MALDI data, which consists of matrix ions, various fragments of matrix ions, matrix clusters such as dimers and trimers, as well as adducts of the matrix molecules with ubiquitous metal ions such as Na^+ and K^+. Software programs have been developed that can be used to predict the mass of the matrix ions that contributes to the low mass background in the spectrum [28].

1.2.1 Electrospray

Dole first introduced the concept of generating gas phase ions from electrically charged liquid droplets in 1968 [29]. Electrospray was coupled to a mass spectrometer (ES-MS), for the detection of biomolecules by John Fenn *et al.* in the mid-1980s [9,30,31]. The electrospray method, displayed in Figure 1.2, creates ions by applying a potential to a flowing liquid causing the liquid to charge and subsequently spray. This spray creates very small droplets of solvent-containing analyte. The electric field at the capillary tip (E_c) can be described mathematically by the following equation:

$$E_c = 2V_c/\{r_c \, ln(4d/r_c)\}, \qquad (1)$$

where V_c is the potential applied to the capillary, r_c is the radius of the capillary, and d is the distance from the counter electrode. At high enough E_c, a liquid filament forms. The liquid filament becomes unstable downstream, which enables the formation of separate droplets. Once these droplets are formed, they undergo shrinkage by evaporation of solvent. This is conventionally accomplished by flowing a drying gas at hot temperatures over the droplets. Evaporation occurs until the Coulombic repulsion forces become sufficient to overcome the surface tension forces. Once the above criteria are met,

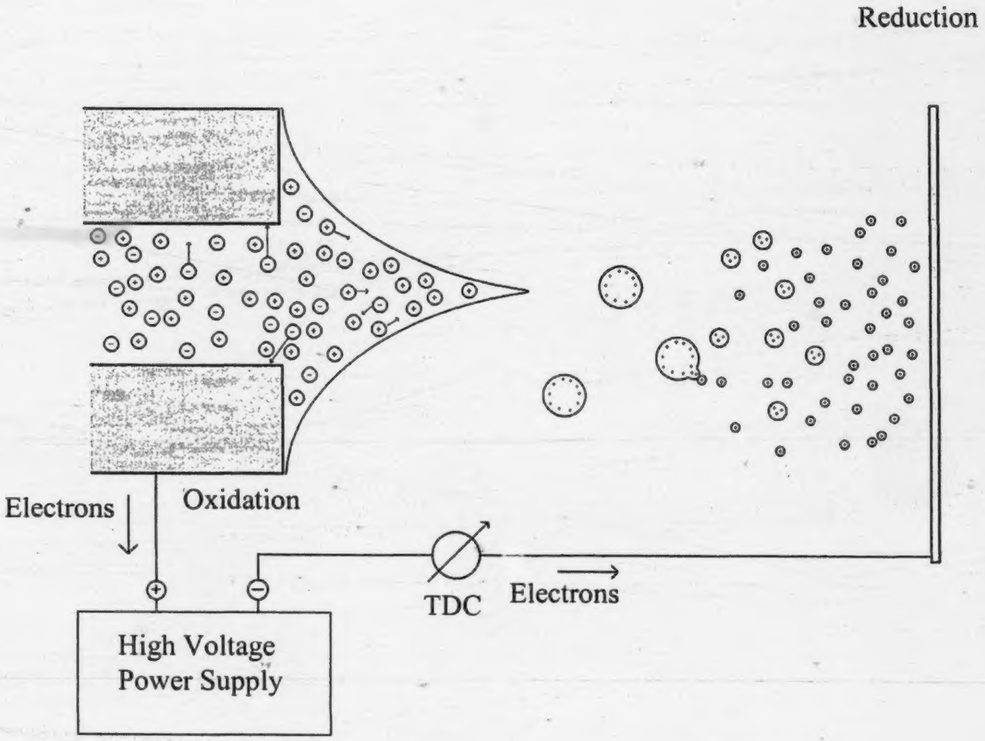

Reduction
Electrons
Oxidation
TDC
Electrons
High Voltage
Power Supply

Figure 1.2 Schematic representation of the electrospray process. First charged droplets are produced at the capillary tip. Solvent evaporation causes shrinkage of the droplets resulting in repeated droplet disintegration until gas phase ions are produced.

the droplet will undergo breakdown to many smaller droplets. This process, know as fission, will repeat itself until the electric field at the liquid surface becomes so high that the solute ions "escape" from the liquid phase to the gas phase (ion-evaporation). Figure 1.2 shows the mechanism of ESI. The ESI process was well described in detail by Kebarle *et al.* [32].

Large ions, such as proteins, typically display a much higher charged state in the electrospray process, as compared to MALDI ionization. The mass spectrometer detects the mass-to-charge ratio of the analytes; therefore high mass proteins with multiply charges are easily detected within a relatively low m/z range [33]. The distribution of charges gives rise to a multiple charge envelope. Figure 1.3 displays the ESI mass spectrum of lysozyme. The spectrum displays several multiple charged ions, up to 13 protons on a single lysozyme molecule, which is typical for large biomolecules. Such a spectrum can be simplified by deconvolution, an algorithm that converts multiple m/z values from a multiple charge envelope into a singly charged peak at the molecular weight of the analyte [34]. Peptides can also be multiply charged, with at +2 or +3 charge states depending on the size and nature of the peptide.

Electrospray ionization is frequently coupled with liquid chromatography for the analysis of complex samples [35]. ESI ionizes the analytes from a solution and is therefore readily coupled to solution-based separation tools. When liquid chromatography and mass spectrometry are coupled (LC-MS), MS analysis of the sample takes place on-line as they elute from the chromatography column.

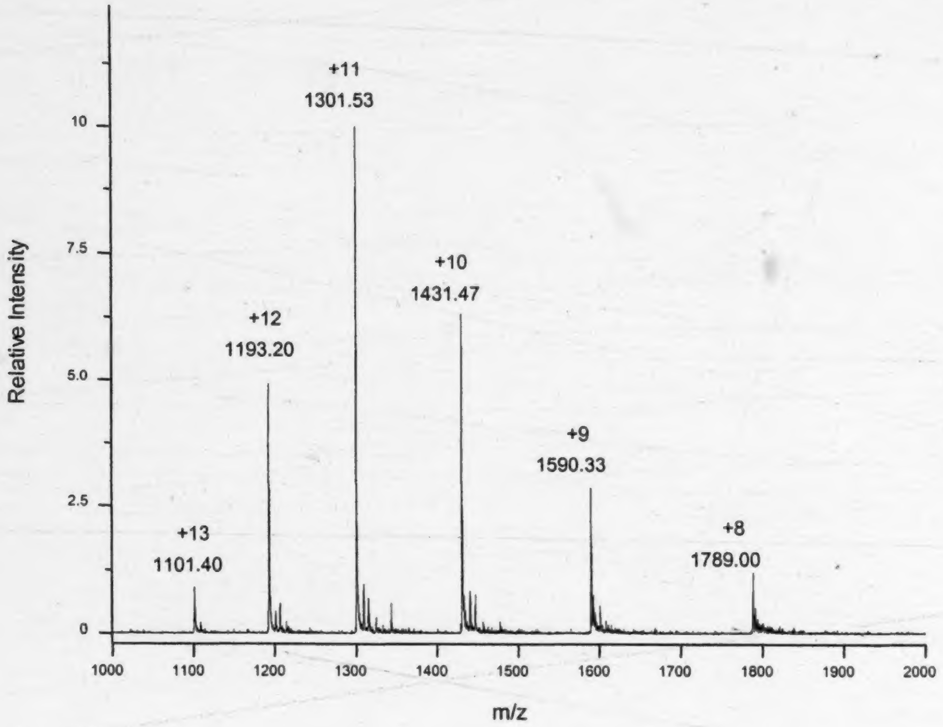


Figure 1.3 ESI mass spectrum of lysozyme. Several multiple charged ions in the spectrum are observed, creating a charged envelope for lysozyme. The charge state of each ion is displayed above its m/z.

1.3 Mass Spectrometers

All mass spectrometers perform three essential functions: ionization, mass analysis, and ion detection. In the first function, an ionization source converts gaseous, solid or liquid phase molecules into gas-phase ions. The second function, performed by the mass analyzer, can be accomplished by one of many different principles. However, the focus of this discussion will be limited to three types, namely separation based on time-of-flight (TOF-MS), separation by quadrupole MS, and separation by selective ejection of ions from a three-dimensional trapping field (ion trap or Fourier transform ion cyclotron MS). The third component is the ion detector. Typical ion detectors include multichannel plates and dynodes. Both detectors involve converting ions to electrons followed by amplification. The magnitude of the current that is produced by the detector as a function of time is used to determine the m/z value of the ion. Both MALDI and electrospray can be coupled to any one of these three methods of mass analysis. The fact that MALDI produces short bursts of ions and electrospray produces a continuous beam of ions typically leads to coupling MALDI with TOF-MS and electrospray with quadrupole or ion-trapping MS. In order to optimize analysis, the method of ionization is based on the nature of the sample and the information required.

1.3.1 MALDI Time-of-Flight Mass Analyzer

In 1946, Stephens first proposed a device that could separate ions in a vacuum tube according to differences in their velocity as they travel through the flight tube [36]. A voltage of 30 kV is typically used to accelerate the ions from the source region through a field free region, known as a drift tube of 0.5 – 2 m in length. The kinetic energy (KE) given to the ions can best be explained with the following equation:

$$KE = zeE\,, \qquad (2)$$

where z is the number of charges on the ion, e is the electron charge, and E is the electric field (30kV). This equation demonstrates that all ions have the same kinetic energy in the source region. Also, the kinetic energy of an ion can be related to its mass, m, according to the equation below:

$$KE = \frac{1}{2} m v^2\,, \qquad (3)$$

where v is the final velocity of the ion. This equation illustrates that once the kinetic energy is imparted, the velocity of the ions will vary if the masses of these ions are different. Combining these two equations allows us to calculate the time required for an ion to travel the flight tube.

$$t = \left(\frac{m}{2zeE}\right)^{\frac{1}{2}} D\,, \qquad (4)$$

where D is the length of the drift tube. By rearranging the above equation, we can correlate the flight time of an ion in a TOF instrument with the mass-to-charge ratio:

$$\frac{m}{z} = 2eE\left(\frac{t}{D}\right)^2. \qquad (5)$$

Therefore, if the length of the flight tube (D) is fixed and the acceleration electric field (E) is constant, then the flight time can be used to calculate the mass-to-charge ratio

There are several advantages to a TOF analyzer: (1) no theoretical mass limit; (2) high mass spectral recording speed; (3) high sensitivity (due to high ion transmission); (4) simplicity and versatility. Since the MALDI ionization process is pulsed, it is well suited for a time-of-flight mass analyzer. Combining MALDI with the theoretically unlimited mass range makes a TOF analyzer ideal for large biomolecule analysis. The MALDI-TOF-MS is an extremely sensitive tool allowing the detection of sample molecules below femtomole levels with mass accuracies generally better than 50 ppm (in a mass range up to about 30 kDa) [37].

1.3.1.1 Time-Lag Focusing

In a MALDI-TOF-MS instrument, the ions produced by the pulsed laser beam are immediately extracted from the source. Therefore there is a constant voltage applied to the source plate. The distribution of ions during the ionization process is the main cause of poor mass resolution in TOF instruments. One technique employed to focus these ions is time-lag-focusing. Figure 1.4 displays a schematic representation of the time lag focusing technology. In the figure, all the ions start at the same position on the MALDI target (1.4A). In a MALDI-TOF instrument equipped with time lag focusing, the ions are first allowed to expand into a field-free region between the repeller and the first extraction grid (1.4B). After a certain time delay, a voltage (E) is applied to the repeller to extract the ions into the flight tube (1.4C). The voltage imparts higher energy to ions closer to the repeller. This energy compensates for the initially lower kinetic energy ions allowing them to reach the detector at the same time as the initially higher energy ions (1.4D). Both the pulse voltage and delay time can be adjusted individually for optimum performance. Whiley and McLaren initially developed time-lag-focusing in 1955 [38]; however, significant improvements have been developed since this technique was introduced [39].

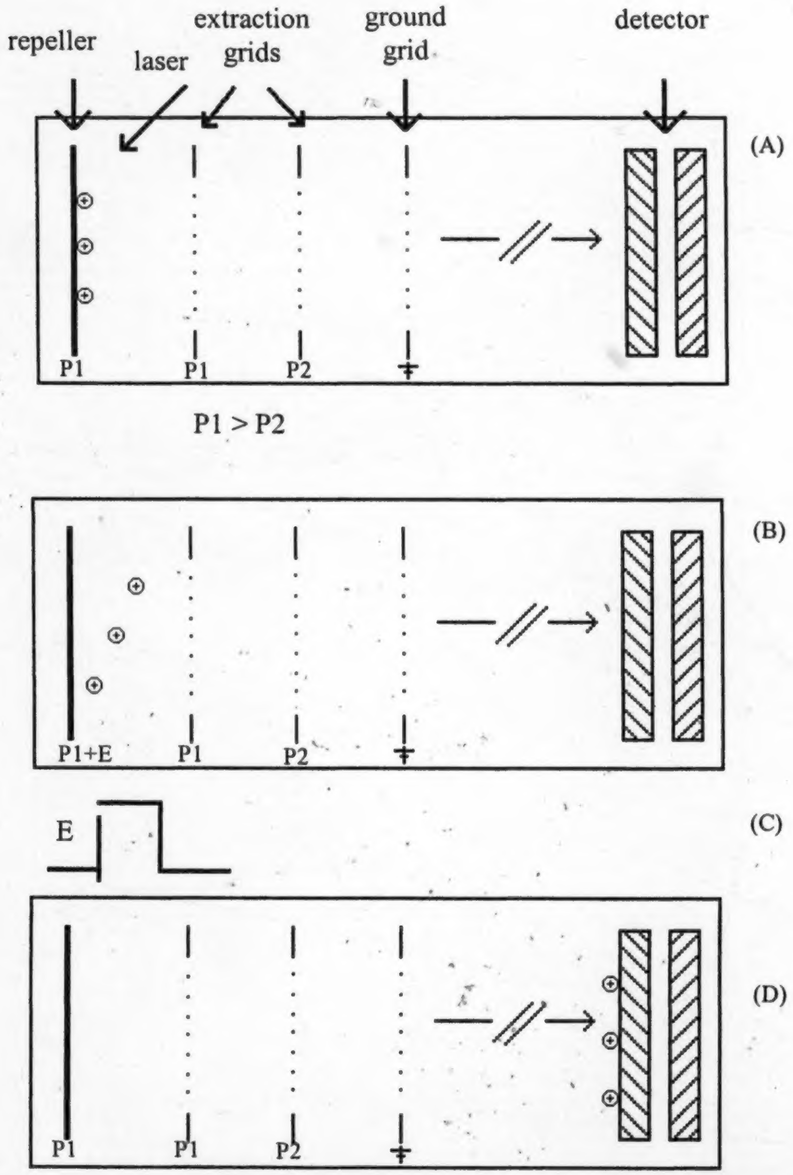

repeller
laser
extraction grids
ground grid
detector
P1
P1
P2
(A)
P1 > P2
P1+E
P1
P2
(B)
E
(C)
P1
P1
P2
(D)

Figure 1.4 Schematic of time-lag-focusing. (A) Ions are desorbed from MALDI target with different initial kinetic energy. (B) Ions are expanded in the field free source region where high energy ions move further away from the repeller. (C) A pulse voltage (E) is applied to compensate for the differenced in initial kinetic energy. (D) Ions reach the detector simultaneously. P1 and P2 are the DC voltages applied on the repeller and extractions grids.

1.3.1.2 Reflectron TOF Instruments

Another technique used to combat spatial distribution of ions in the source region is to place an ion mirror, or reflector, at the end of the flight tube. This technique was developed by Mamyrin in 1973 [40]. Figure 1.5 displays a schematic diagram of a reflectron TOF instrument. The ion reflector consists of a series of ring electrodes. The voltage applied to each electrode increases along the length of the flight tube. Ions travel down the flight tube, encounter the repulsion of the ion mirror, and hit a detector at the source end of the flight tube. From Figure 1.5, ion (A) started with higher kinetic energy and therefore will travel faster and penetrate further into the ion mirror, therefore it will travel a longer distance than ion (B) before reversing direction, and take longer to hit the detector than slower ions. Consequently, fast (A) and slow (B) ions of the same mass to charge ratio are focused in time at the detector, thus improving resolution. Resolution of 37,000 (FWHM) has been recorded for proteins with modern reflectron TOF instruments [41].

1.3.2 Quadrupole Mass Analyzer

Electrospray is most often combined with a quadrupole mass spectrometer. The quadrupole is a mass filter, consisting of four cylindrical rods to create a hyperbolic field. The same absolute potential ($U + V\cos(\omega t)$) with different polarity is applied to each rod, where U is DC voltage, V is AC or RF voltage, ω is the angular velocity, and t is time. When the rod voltage is positive, the positive ions are repulsed, and are pushed towards the center (focus) of the four rods. The reverse is true when a negative voltage is applied (defocus). Therefore, the rod potentials are set so that only ions within a small m/z range

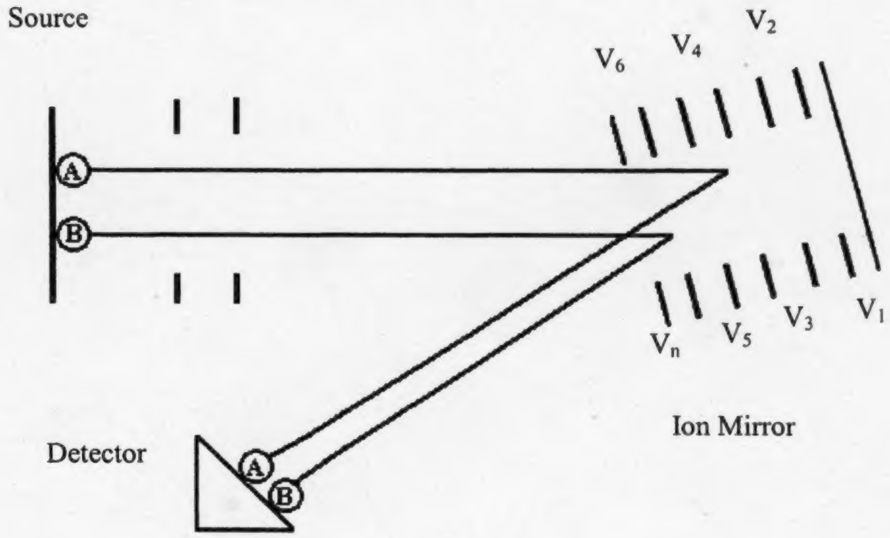

Source
V$_6$
V$_4$
V$_2$
A
B
V$_n$
V$_5$
V$_3$
V$_1$
Ion Mirror
Detector
A
B

Figure 1.5 Schematic diagram of a reflectron time-of-flight mass spectrometer. V is the voltage applied to each plate in the ion mirror. $V_1 > V_2 > V_3 > V_n$.

have a stable trajectory and are transmitted to the detector. All other ions will collide with rods and are therefore not transmitted. If the voltages V and U are simultaneously increased, while keeping their ratio constant, ions are transmitted successfully from low to high mass.

1.3.3 Quadrupole IonTrap

A quadrupole ion trap consists of two end cap electrodes and a ring electrode (displayed in Figure 1.6). Appling a potential on the electrodes creates a hyperbolic electric field inside the chamber. Ions can be created by electrospray ionization and transmitted into the ion trap. The ions are then focused near the centre of the ion trap by applying the appropriate RF voltage. The quadrupole ion trap is capable of two modes of detection. The entire m/z ranging, from 100 up to 2500, can be detected. The trap is also capable of selecting one ion with a particular m/z value. This ion can then undergo collisions with He and the subsequent fragments can be detected. In this mode, termed tandem mass spectrometry or MS/MS, the voltages of the ion trap are selected to eliminate all but the desired m/z ion(s) in preparation for subsequent mass analysis. The RF voltages applied to the ion trap are adjusted to stabilize the desired ions and capture any ions formed by collisions. The desired ions then are given energy by application of supplemental voltage to the end caps to cause dissociating collisions with the dampening gas (He) and are retained in the trapping field. Once the fragmentation is completed, scanning is accomplished by changing the RF voltage applied in the ring electrode of the ion trap. This change moves the ions in the stability diagram to a position where they become unstable in the z-direction and leave the trapping field for detection. Ions are

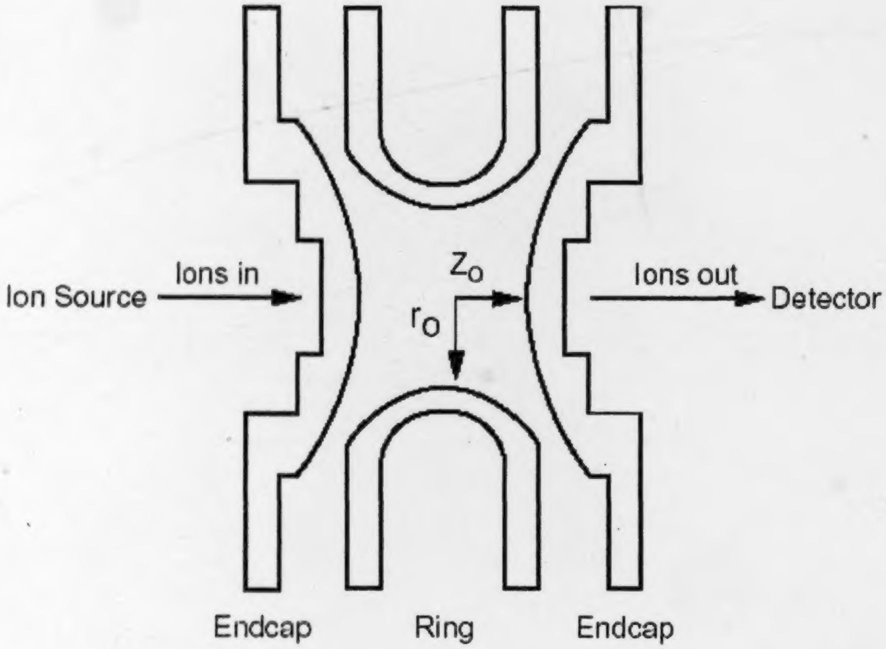

Ion Source
Ions in
Z_o
r_o
Ions out
Detector
Endcap
Ring
Endcap

Figure 1.6 Schematic of a quadrupole ion trap mass spectrometer. The ion trap collects ions and then can be used to select an ion of interest and discard the rest. The ion of interest can then be fragmented by introducing a collision gas (He). The fragmented ions are then ejected out of the trap for detection.

accelerated into a high voltage dynode where they are converted into electrons that are amplified by an electron multiplier. The procedure for performing MS^2 or MS^n (where n>2) is described in detail by Cooks *et al.* [42]. This technology is particularly beneficial for structural analysis of peptides.

1.4 Protein Identification

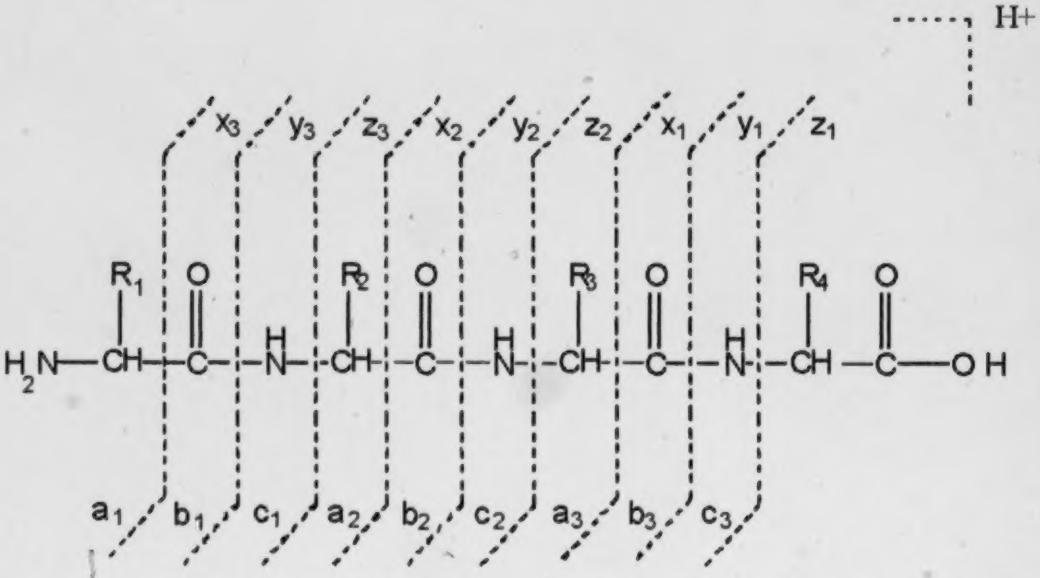
As previously discussed, the process of sequencing a protein prior to the development of mass spectrometry for protein sequencing was an extremely laborious process. With the use of mass spectrometry, sequencing proteins become relatively easier. Electrospray is the ionization method of choice for determining the precise molecular weight of proteins. However, if an approximate molecular weight is sufficient, MALDI may be used. MALDI results in broad peaks and low sensitivity for proteins above 30 kDa. However, the intact molecular mass of a protein is typically not enough to identify an unknown. Two techniques have gained widespread use to gain more information: peptide mass mapping and tandem mass spectrometry.

1.4.1 Peptide Mass Fingerprinting

In the technique of peptide mass fingerprinting (PMF), in order to identify a protein, it must first be broken down into several small pieces (peptides). In PMF, an enzyme is often used to degrade the protein of interest. Trypsin is most frequently used, being a well characterized enzyme that selectively cleaves on the carboxy-terminal side of arginine and lysine. The enzyme specificity creates predictable peptides allowing for comparison with theoretical databases. The digestion usually results in the production of several peptide fragments of the protein. The masses of these peptides are then accurately determined by mass spectrometry. The obtained peptide masses are then compared to a list of theoretically expected tryptic peptide masses for each entry in a database. Known protein sequences are used to create the database. Some examples of protein database searching programs are MS-FIT (http://prospector.ucsf.edu/ucsfhtml4.0/msfit.htm), PeptIdent (http://us.expasy.org/tools/peptident.html), and Mascot (www.matrixscience.com). The possible identities of the original protein are then ranked based on the number of peptides matching. In general, more peptides from a protein searched against the database results in higher confidence level for its identification. The ability to acquire highly accurate m/z values has helped this method of protein identification a great deal. As the accuracy of molecular measurements increases, the number of peptides that will match to those in the database will decrease. This results in increasing the confidence of the identification. The method of peptide mass fingerprinting was simultaneously published by four independent research groups in 1993 [43-47]. This method is commonly employed for protein identification; however, it is limited to relatively pure samples. If there are several proteins present, the mass spectrum becomes very complicated and increasing the difficulty to discern which peptide came from which protein.

1.4.2 MS Fragmentation of Peptides

A second method for protein identification employs peptide fragmentation data that are generated by MS/MS. In the collision-induced dissociation (CID) process,



H+
x3
y3
z3
x2
y2
z2
x1
y1
z1
R1
R2
R3
R4
O
O
O
O
H
H
H
H2N—CH—C—N—CH—C—N—CH—C—N—CH—C—OH
a1
b1
c1
a2
b2
c2
a3
b3
c3

Figure 1.7 Nomenclature system for fragment ions generated by collision-induced dissociation of peptides. The ions labeled a, b, and c refer to charge retention on the N terminus, whereas x, y, and z ions retain the charge on the C terminus.

peptide ions fragment in a predictable manner. Figure 1.7 shows the nomenclature system for peptide fragments generated by collision-induced dissociation [48]. The system classifies ions based on where the bond breakage occurs and whether the charge is carried on the carboxy-terminal or the amino-terminal side of the peptide. Thus, sequences from the database can be used to predict an expected fragmentation pattern and match the expected pattern to that observed in the spectrum.

Traditionally, peptide sequencing was performed by ESI-MS. However, recently MALDI-MS instruments capable of peptide fragmentation and detection have become available. In the ESI-MS method, the ionization technique generally produces multiply charged peptides. Tryptic peptides are predominately doubly or triply charged.

Most proteomics experiments are carried out with tryptic peptides, which have arginine or lysine as their C-terminal residues. In low energy CID, as seen in ion trap mass spectrometry, typical fragmentation patterns would be dominated by the b- and y-ions. The b- and y-ions are a result of the amide bond fragmenting with the charge retention on the N or C terminus respectively. An advantage of this approach is that it provides much more information than the intact molecular weight of the peptide. Each peptide MS/MS spectrum represents a unique piece of information; consequently, matching one or more MS/MS spectra to a sequence of the same protein provides a high level of confidence in the identification.

Because the instrument can isolate a peptide very accurately from a collection of other ions, it provides a powerful tool to analyze peptide mixtures. When numerous peptide ions enter an ion trap, for example, one peptide m/z value can be isolated and then fragmented to obtain a fragmentation pattern that is unique to an amino acid sequence. The fragmentation m/z values are searched against a database of known protein sequences. Matching a tandem mass spectrum uniquely to a sequence allows proteins to be identified.

1.4.3 Database Searching

Genome sequencing or expressed sequence tags (EST) have stimulated the development of databases allowing identification of proteins through extensive searching algorithms [49]. In recent years, large steps have been made in developing software to interpret large quantities of data generated by mass spectrometry [50,51]. These programs have the ability to identify proteins *via* various scoring systems. Currently, rapid and automated protein identification can be achieved by searching nucleotide or protein sequence databases. Of the search programs available for protein identification by tandem mass spectrometry, SEQUEST, developed by J. Yates and J. Eng at the University of Washington, is one of the most common software algorithms used [52]. This program searches experimental peptide fragments *vs*. a theoretical database or subset database of proteins. The possible sequences are ranked based on the number of ions matching and the uniqueness of the match. There are several parameters used to analyze the difference between the first and second-ranked sequences. Appendix A summarizes each parameter. This software enables the analysis of hundreds of tandem mass spectra in approximately 1 hour, depending on the size of the database searched. Tandem mass spectrometry technology used to identify proteins has significantly increased the confidence of protein identification and offers a complementary technique to peptide mass fingerprinting.

1.5 Sample Preparation and Analysis by Mass Spectrometry

1.5.1 Difficulties in Sample Preparation Prior to Mass Spectrometry

1.5.1.1 Sample Contamination

Mass Spectrometry is the method of choice for identification and characterization of proteins from various organisms [53-56]. There are different mass spectrometric strategies for protein identification. However, these methods are not yet considered routine. Sample preparation prior to mass spectrometry is the key to successful identification and characterization of proteins. Protein samples require several conditioning steps prior to analysis. Procedures for extracting proteins from biological matrices often involve the addition of buffers, salts, or surfactants to the sample. Also, the sample may contain other components (lipids, carbohydrates, and DNA) along with the protein(s) of interest. MALDI is known for its tolerance, to a certain degree, of contaminants such as non-volatile salts [57]. However, when samples containing high amounts of salts and/or buffers are analyzed by MALDI-MS, the contaminants can contribute to a loss in signal intensity, or complete signal suppression. ESI has an even lower tolerance for contamination. Thus, biological samples require purification strategies prior to MS analysis.

Several methods are available for sample clean up prior to analysis including ion exchange chromatography, reversed-phase chromatography, dialysis, precipitation, and solvent extraction. These strategies are subjected to loss of protein during the sample manipulation steps, thus jeopardizing the amount of protein remaining for analysis. Also, in some techniques, unwanted salts or buffers are needed for optimal performance. For example, while ion chromatography may separate the proteins, the method introduces contamination through the addition of salt. Therefore, multiple clean up and concentration steps are typically employed. This could involve the used of commercially available microcolumns such as Ziptips. These small cartridges contain reversed-phase resins for selective removal of contaminants with minimal loss of sample [58].

1.5.1.2 Protein Digestion

At present, it is the protein digestion step that limits the ultimate sensitivity of the MS technique for protein identification. Specifically, protein digestion at submicromolar concentration is very difficult. As the protein concentration is reduced, the kinetics of the digestion reaction is reduced to the point where impractical digestion times are required. The enzyme-substrate reaction rate, ν, to a first approximation, is expressed below:

$$\nu = \frac{k_{cat}[E]_0[S]}{K+[S]}, \qquad (6)$$

where $[E]_0$ is the total concentration of enzyme, [S] is the concentration of substrate (the protein), K is the equilibrium constant of the enzyme-substrate complex, and k_{cat} is the turnover rate constant for the formation of product. To avoid these long digestion times, one can increase the concentration of the digesting enzyme $[E]_0$. However, the presence of a high enzyme-to-sample ratio often produces a high degree of enzyme autolysis. These enzyme fragments can easily mask the signals of the actual sample peptides in the MALDI or ESI spectrum. Thus, low concentrations of enzyme are used when digesting an extremely low concentration of protein. This results in slow reaction rates and limits the concentration of sample required for analysis. Several procedures exist for protein degradation including immobilized enzymes cartridges [59], direct on-probe digestion [60] and digestion in minute sample volumes [61].

1.5.2 Sample Preparation Automation

Currently, two-dimensional gel electrophoresis (2DE) is the method of choice for separation complex cell lysates [6,62]. However, 2DE is limited by extensive sample handling, non-linear response factors for the most commonly used staining techniques, limited loading capacity, low extraction efficiencies of the gel-embedded proteins, and a decreasing resolving power for proteins with a molecular mass of less than 15 kDa. Furthermore, it is difficult to isolate proteins at acidic and basic pH extremes, proteins with a molecular weight higher than 200 kDa, or membrane and hydrophobic proteins [63]. In addition, this method is time-consuming and difficult to automate. Therefore several methods exist as potential alternatives to 2D gels [64-66].

High performance liquid chromatography (HPLC) is one such technique that offers a potential alternative to 2D gels. Single dimension HPLC typically does not have the chromatographic resolution for proteome analysis of complex biological systems [67]. Thus, multi-dimensional liquid chromatography is used. This method involves coupling two or more different modes of separation together. This technique has been successfully automated and directly coupled to mass spectrometry. Techniques based in multi-dimensional HPLC currently lead the field of proteomics for high throughput protein identification. There are several combinations of techniques used including reversed-phase chromatography, ion exchange chromatography, size exclusion chromatography, and affinity chromatography.

1.6 Challenges Associated with Proteome Analysis

With the innovation of soft ionization techniques, mass spectrometry becomes the tool of choice for protein identification and has lead to the exponential growth of proteomics. Mass spectrometry enables the rapid identification of low abundant proteins. Currently, the goal of proteomics is to rapidly identify all of the proteins expressed by a cell or tissue. Despite all the advances in mass spectrometry and separations, such an undertaking is not possible with existing methods. Proteomics must deal with unavoidable problems such as the high degree of complexity of cellular proteomes, limited and variable sample material, sample degradation, vast dynamic range (more than 10^6-fold for protein abundance), and post-translation modifications [68-70].

The analysis of a complete proteome has yet to be achieved. Yates *et al.* have identified the largest number of proteins in a single experiment to date with 1,484 from yeast; however, there are a predicted 4000 – 6000 proteins present. It is therefore evident that even the most state-of-the-art technologies do not address the requirements for proteome analysis.

In this thesis, a method is described to concentrate, clean, and digest biological samples for the purpose of protein identification. An extensive study on the digestion characteristic of surface bound proteins will be explored. Further understanding of surface digestion will then lead to designing a system capable of high throughput analysis. This online automated system will be optimized and *E. coli* whole cell lysate will be analyzed. The system combines sample manipulation and peptide sequencing, providing a tool for potentially high throughput protein identification.

1.7 Literature Cited

1. Wilkins, M. R.; Sanchez, J. C.; Gooley, A. A.; Appel, R. D.; Humphery-Smith, I.; Hochstrasser, D. F.; Willians, K. L. *Biotechnol. Genetic Eng. Reviews* **1995**, *13*, 19-50.
2. Sanger, F. *J. Biochem.* **1945**, *39*, 507.
3. Ryle, A. P.; Sanger, F. *J. Biochem*, **1955**, *60*, 535.
4. Pauling, L.; Itano, H. A.; Singer, S. J.; Wells I. C. *Science* **1949**, *110*, 543.
5. Ingram, V. M. *Nature* **1956**, *178*, 792-794.
6. O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007.
7. Edman, P. *Acta Chem. Scand.* **1953**, *7*, 700–701.
8. Aebersold, R. H.; Pipes, G.; Hood, L. E.; Kent, S. B. *Electrophoresis* **1988**, *9*, 520-530.
9. Mann, M.; Hendrickson, R. C.; Pandey, A. *Ann. Rev. Biochem.* **2001**, *70*, 437-473.
10. Fenn, J. B.; Mann, M.; Meng, C.K.; Wong, S.F. *Science* **1989**, *246*, 64–71.
11. Tanaka, K.; Ido, Y.; Akita, S. *Proceedings of the Second Japan-China Joint Symposium on Mass Spectrometry*; Matsuda, H., Liang, X.-T., Eds.; Bando Press: Osaka, Japan, **1987**, 185-188.

12. Tanaka, K.; Waki, H.; Ido, Y.; Akita, S.; Yoshida, Y.; Yoshida, T. *Rapid Commun. Mass Spectrom.* **1988**, *2*, 151-153.
13. Karas, M.; Bachmann, D.; Bahr, Y.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Processes* **1987**, *78*, 53-68.
14. Karas, M.; Hillenkamp, F. *Anal. Chem.* **1988**, *60*, 2299-2301.
15. Hillenkamp, F.; Karas, M. *Int. J. Mass Spectrom.* **2000**, *200*, 71-77.
16. Karas, M.; Bachmann, D.; Hillenkamp, F. *Anal. Chem.* **1985**, *57*, 2935-2939.
17. Hillenkamp, F.; Karas, M.; Holtkamp, D.; Klüsener, P. *Int. J. Mass. Spectrom. Ion Process.* **1986**, *69*, 265-276.
18. Schreimer, D. C.; Li, L. *Anal. Chem.* **1996**, *68*, 2721-2725.
19. Karas, M.; Bahr, U.; Gieβmann, U. *Mass Spectrom. Rev.* **1991**, *10*, 335-357.
20. Beavis, R. C.; Chait, B. T. *Chem. Phys. Letters* **1993**, *181*, 470-484.
21. Karas, M.; Glückmann, M.; Schäfer, J. *J. Mass Spectrom.* **2000**, *35*, 1-12.
22. Kinsel, G.; Edmondson, R. D.; Russel, D. *J. Mass. Spectrom.* **1997**, *32*, 714-722.
23. Karas, M.; Hillenkamp, F. *Organic Mass Spectrom.* **1992**, *27*, 472-480.
24. Zenobi, R.; Knochenmuss, R. *Mass Spectrom. Rev.* **1998**, *17*, 337-366.
25. Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087-1091.
26. Krause, E.; Wenschuh, H.; Jungblut, P. R. *Anal. Chem.* **1999**, *71*, 4160–4165.
27. Dai, Y. Q.; Whittal, R. M.; Li, L. *Anal. Chem*. **1996**, *68*, 2721-2725.
28. Keller, B. O.; Li, L. *J. Am. Soc. Mass. Spectrom.* **2000**, *11*, 88-93.
29. Dole, M. J. *J. Chem. Phys*, **1968**, *49*, 2240.
30. Yamashita, M.; Fenn, J. B. *J. Phys. Chem*. **1984**, *88*, 4671.

31. Whitehouse, C. M.; Dreyer, R. N.; Yamashita, M.; Fenn, J. B. *Anal. Chem.* **1985**, *57*, 675.

32. Kebarle, P.; Tang, L. *Anal. Chem.* **1993**, *65*, 972A.

33. Loo, J. A.; Udseth, H. R.; Smith, R.D. *Anal Biochem.* **1989**, *179*, 404-412.

34. Barbalas, M. P.; Garland, W. A. *J Pharm Sci.* **1991**, *10*, 922-927.

35. Brudel, M.; Kertscher, U.; Schroder, D.; Melzig, M. F.; Mehlis, B. *J. Chromatogr. A* **1995**, *712*, 169-75.

36. Stephens, W. E. *Phys. Rev.* **1946**, *69*, 691.

37. Wurz, P.; Gubler, L. *Rev. Sci. Instrum.* **1996**, *67*, 1790-1793.

38. Wiley, W. C.; McLaren, I. *Rev. Sci. Instrum.* **1955**, *26*, 115-1157.

39. Whittal, R.; Russon, L.; Weinberger, S. R.; Li, L. *Anal. Chem.* **1997**, *69*, 2147-2153.

40. Mamyrin, B. A.; Karatajev, V. J.; Shmikk, D. V.; Zagulin, V. A. *Soviet Phys. JETP* **1973**, *37*, 45-48.

41. Vestal, M. L. *Chem. Rev.* **2001**, *101*, 361-375.

42. Louris, J. N.; Cooks, R. G.; Syka, J. E. P.; Kelley, P. E.; Stafford, G. C.; Todd, J. F. J. *Anal. Chem* **1987**, 59, 1677-1685.

43. Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 5011-5015.

44. Mann, M.; Hojrup, P.; Roepstorff, P. *Biol. Mass Spectrom.* **1993**, *22*, 338-345.

45. Yates III, J. R.; Speicher, S.; Griffin, P. R.; Hunkapiller, T. *Anal. Biochem.* **1993**, *214*, 397-408.

46. James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G. *Biochem. Biophys. Res. Comm.* **1993**, *195*, 58-64.

47. Pappin, D. J. C.; Hojrup, P.; Bleasby, A. *J. Curr. Biol.* **1993**, *3*, 327-332.

48. Biemann, K.; Scoble, H. A. *Science*, **1987**, *237*, 992-998.

49. Adams, M. D.; Dubnick, M.; Kerlavage, A. R.; Moreno, R.; Kelley, J. M.; Utterback, T. R.; Nagle, J. W.; Fields, C.; Venter, J. C. *Nature* **1992**, *270*, 467-470.

50. Yates III, J. R.; McCormack, A.L.; Schieltz, D.; Carmack, E.; Link, A. *J. Protein Chem.* **1997**, *16*, 495–97.

51. Fenyo, D.; Beavis, R. C. *Trends in Biotech*. **2002**, *20*, S35-S38.

52. Ducret, A.; Van Oostveen, I.; Eng, J. K.; Yates III, J. R.; Aebersold, R. *Protein Sci.* **1998**, *7*, 706-719.

53. Link, A. J.; Hays, L. G.; Carmack, E. B.; Yates III, J. R. *Electrophoresis* **1997**, *18*, 1314-1334.

54. Shevchenko, A.; Jenson, O. N.; Podtelejnikov, A. V.; Sagliocco, F.; Wilm, M.; Vorm, O.; Mortensen, P.; Boucherie, H. Mann, M. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 1440-1445.

55. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E., Mize, G. J.; Morris, D. R. Garvik, B. M, Yates III J. R. *Nat. Biotechnol.* **1999**, *17*, 676-682.

56. Gygi, S. P., Rist, B., Gerber, S. A.; Turecek, F. Gelb, M. H., Aebersold, R. *Nat. Biotechnol.* **1999**, *17*, 994-999.

57. Yates III, J. R. *J. Mass Spectrom.* **1998**, *33*, 1-19.

58. Jin, X.; Chen, Y.; Lubman, D. M.; Misek, D.; Hanash, S. M. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 2327-2334.

59. Hsieh, Y. L.; Wang, H.; Elicone, C.; Mark, J.; Martin, S. A.; Regnier, F. *Anal. Chem.* **1996**, *68*, 455-462.

60. Owens, D. R.; Bothner, B.; Phung, Q.; Harris, K.; Siuzdak, G. *Bioorganic Med. Chem.* **1998**, *6*, 1547-1554.

61. Whittal, R. M.; Keller, B. O.; Li, L. *Anal. Chem.* **1998**, *70*, 5344-5347.

62. Lilley, K. S.; Razzaq, A.; Dupree, P. *Curr. Opin. Chem. Biol.* **2002**, *6*, 46-50.

63. Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C. *Electrophoresis* **2001**, *21*, 1104.

64. Patterson, S. D. *Curr Opin Biotechnol.* **2000**, *11*, 413-418.

65. Hong, W.; Hanash, S. *J. Chromatogr. B* **2003**, *787*, 11-18.

66. Hille, J. M.; Freed, A. L.; Watzig, H. *Electrophoresis.* **2001**, *22*, 4035-4052.

67. Davis, J. M.; Giddings, J.C. *Anal Chem.* **1983**, *55*, 418-424.

68. Kellner, R. *Fres J. Anal. Chem.* **2000**, *366*, 517-524.

69. Godovac-Zimmermann, J.; Brown, L. R. *Mass Spectrom. Rev.* **2001**, *20*, 1-57.

70. Larsen, M. R.; Roepstorff, P. *Fres J. Anal. Chem.* **2000**, *366*, 677-690.

Chapter 2

Microcolumn Capture and Digestion of Proteins Combined with Mass Spectrometry for Protein Identification [1, 2]

2.1 Introduction

Protein identification and characterization is routinely accomplished with mass spectrometry-based approaches. A fundamental step in most MS approaches for protein characterization involves the controlled degradation of protein samples to form fragments of the protein, prior to MS analysis. This is typically accomplished through enzymatic hydrolysis of the protein sample. For example, peptide mass fingerprinting, one of the key advancements in rapid protein identification [1-4], combines sequence-specific enzyme digestion of the protein with MS analysis of the resulting peptides, most typically with matrix-assisted laser desorption/ionization (MALDI) time-of-flight (TOF) MS, as a

---

[1] A form of this chapter has been published as: D. Craft, A. Doucette and Liang Li "Microcolumn Capture and Digestion of Proteins Combined with Mass Spectrometry for Protein Identification" *Journal of Proteome Research*, **2002**, *1*, 537-547.

[2] A. Doucette made significant contributions towards the analysis of human plasma and tears.

means of generating data for protein identification. Likewise, the complementary technique of peptide amino acid sequencing through tandem MS fragmentation by collision-induced dissociation or post-source decay, also incorporate enzyme digestion of the protein [1-5]. In this case, the generation of smaller segments of the protein is desired since these lower-molecular weight segments are more easily fragmented and detected by the MS instrumentation. Effective digestion of the protein sample is therefore vital in achieving successful MS characterization of protein samples.

Although the enzymatic digestion of proteins is conceivably a simple process, the complexity of biological protein samples leads to several potential problems. Of particular importance is the high dynamic range for proteins in cells (between 10^5 and 10^6) [6]. Many proteins are naturally expressed at low abundance. Furthermore, difficulties associated with the extraction or solubilization of proteins during sample workup may also lead to the recovery of only small quantities of analyte for digestion. In many cases, proteins are isolated at relatively low concentrations (sub-micromolar) in a relatively large volume (hundreds of microliters or more) [7]. The digestion of proteins at low concentrations is inherently difficult, since most of the commonly used endopeptidases and exopeptidases have K_m values on the order of 5 to 50 mM [8]. Digestion efficiency is therefore significantly hindered as the concentration of protein is reduced below micromolar levels. To compensate for this, the reaction time for the digestion may be increased. Alternatively, the concentration of enzyme may be raised. However, raising the enzyme concentration relative to the amount of protein in the sample has the undesirable effect of increasing enzyme autolysis, which can results in suppression of MS signals for the analyte.

Many procedures have been developed for the digestion and subsequent MS analysis of small quantities of protein, including in-gel digestion of low amounts of sample [9], digestion of small volumes of proteins in enclosed microfluidic devices [10,11], or in nanovials [12-14], and in-capillary digestion of nanoliter volumes of sample [15]. These methods have proven useful for the analysis of extremely low starting quantities of sample. Other techniques focus on increasing the sample concentration or on improving the digestion kinetics. For example, sample preconcentration techniques can be employed prior to enzyme digestion. This could involve solvent evaporation, membrane cutoff filtration, protein affinity capture, or binding to a non-specific stationary support. The use of immobilized enzymes in cartridges or on membranes has frequently been used as a means of increasing the enzyme concentration relative to the protein sample, while minimizing enzyme autolysis [16-21]. In previous work [22], we reported on a method for digestion of dilute protein samples. The method incorporates a preconcentration step prior to digestion by adsorption of the protein onto reversed-phase chromatography beads, placed directly into the sample solution contained within vials. The protein sample is then digested while still adsorbed to the surface. Detection of the resulting peptides was accomplished using MALDI-TOF MS, by directly placing the beads on a MALDI target for analysis. The adsorption step served a dual purpose in allowing for sample preconcentration, as well as sample cleanup prior to digestion. However, the method suffers from difficulties associated with the manipulation of beads in the vials, particularly for automated analysis. Also, the detection method was limited to MALDI analysis in that it is difficult to elute the generated peptides from the bead surface into a small volume of sample for subsequent analysis by other detection means.

In this chapter, a method is presented for the digestion of proteins while adsorbed to a hydrophobic support contained in a microcolumn. This microcolumn protein capture and digestion method offers the potential for automation for parallel sample processing, as well as for incorporation into an integrated system that allows for protein capture, sample cleanup, and digestion, with on-line analysis of the digestion products by mass spectrometry. The performance of this procedure is demonstrated through in-column digests of various protein samples and protein mixtures.

2.2 Experimental

2.2.1 Materials and Reagents

The protein standards used in this study, including the enzyme trypsin (from bovine pancreas, TPCK treated to reduce chymotrypsin activity; dialyzed, lyophilized) were from Sigma Aldrich Canada (Ontario, Canada). Human plasma was the gift of a male healthy donor, and was fractionated according to the Cohn precipitation method, using 1 mL of plasma as starting material [23]. Fraction #IV-4 from the Cohn precipitates was used in this study. BL21(DE3)pLysS *E. coli* cells, cloned to express cardiac Torponin C (cTnC) [24], were obtained from Monica Li in the Department of Biochemistry, University of Alberta. Poros 20 R2 beads (20 μm) were from PerSeptive Biosystems (a gift of Professor F. Cantwell, University of Alberta). The 1-200 µL Microcapillary Tips (gel loader tips) were purchased from Rose Scientific (Alberta, Canada). Glass microfiber filters were purchased from Whatman (Maidstone, UK). Analytical grade acetone, methanol, acetonitrile, acetic acid, and trifluoroacetic acid (TFA) were purchased from Caledon Laboratories (Alberta, Canada). Water used in the experiments was from a NANOpure water system from Barnstead/Thermolyne (Dubuque, Iowa, USA). All other chemicals were purchased from Sigma Aldrich Canada.

2.2.2 *E. coli* Protein Extraction

The *E. coli* cells were amplified in 2×LB media containing ampicillin (100 mg/mL) and chloramphenicol (30 mg/mL) until an O.D. of 0.5 was reached, followed by isopropylthio-beta-galactoside (IPTG) induction to a concentration of 0.4 mM IPTG, and incubation at 37°C for 4 hours. The cells were then harvested by centrifugation and lyophilized. The protein extract was prepared using a previously described method [25]. Briefly, the cells were suspended in 50 mM Tris, at a concentration of 2 mg/mL. The suspension was vortexed for ~15 min, centrifuged, and 1 mL of the clear supernatant was collected. The supernatant was filtered through a Microcon-3 filter from Millipore (Billerica, Massachusetts, USA) with a molecular weight cutoff of 3000 Da. The final collection volume was ~100 µL.

2.2.3 Anion Exchange Chromatography

The *E. coli* extract was separated by anion exchange chromatography. The column used was a 4.6 × 50 mm Mono Q HR 5/5 column from Amersham Pharmacia

Biotech AB (Upsalla, Sweden). The anion exchange beads were 10 μm in diameter with a column capacity of 50 mg. The solvents used were 20 mM Tris in water, buffered to a pH of 8 with HCl (A), and 20 mM Tris in water, pH 8, containing 1 M NaCl (B). The gradient used for the separation was as follows: Solvent B was held at 0% for 5 min, then increased to 35% over 25 min, then to 50% in the following 10 min. A flow rate of 1 mL/min was used, with fractions collected in one-minute intervals. A 100-μL aliquot of the filtered *E. coli* extract was used for the separation.

2.2.4 Protein Digestion in Microcolumns

The design of the microcolumn is based on the work presented by Roepstorff *et al.* [26] and by Annan *et al.* [27]. Briefly, an aliquot of the packing material, suspended in methanol at a concentration of 10 mg/mL, was transferred to a microcapillary gel loader tip (1-200 μL), which had been pinched at the end with pliers. A small piece of glass microfiber membrane, placed at the end of the microcapillary tip, was used as a frit to ensure that the microbeads did not pass through the tips. The procedure for sample loading and digestion is based on previous work [28]. Briefly, a 100 μL protein sample, at a concentration of 100 nM, was pumped through the column a total of three times at an approximate flow rate of 10 μL/min, thus allowing 30 minutes for sample loading. The column was then washed with 100 μL of water. Lysozyme was reduced in-column to break disulfide bonds by flowing 10 μL of 25 mM dithiothreitol (DTT) through the column, followed by incubation at room temperature for 20 minutes. The cysteines were then blocked to prevent reformation of disulfide bonds using 10 μL of 50 mM iodoacetamide, with incubation at room temperature in the dark for 20 minutes. The absorbed protein was digested by flowing a 10 µL solution of 31-ng/µL trypsin in 50 mM NH_4HCO_3 through the column. A small volume of enzyme solution was left in the column to prevent the beads from drying and aid in digestion. The column was incubated for 30 minutes at 37°C and, following digestion, the sample was eluted into a vial using 3 µL of 50% (v/v) acetonitrile/ 0.1% aqueous TFA.

2.2.5 MALDI Sample Preparation

The matrix α-cyano-4-hydroxycinnamic acid (HCCA) was used in these experiments. HCCA was first purified by recrystallization from ethanol. For the column digests, a 0.4 µL portion of undissolved HCCA suspended in 50% (v/v) acetonitrile/ 0.1% TFA was added to the 3 µL column extract; thus ensuring saturation of the sample with matrix. The sample-matrix suspension was vortexed briefly and then centrifuged for 2 min to settle any undissolved matrix. A 0.7 µL portion of the solution was deposited on a MALDI target and allowed to dry in air. On-probe washing was performed by placing a 1 µL droplet of room temperature water on the dried sample spot, then blowing the water droplet off with pressurized air after a few seconds.

2.2.6 Mass Spectrometry

MALDI spectra were collected on a linear time-lag focusing MALDI TOF mass spectrometer, which was constructed at the University of Alberta and has been described in detail elsewhere [29]. A pulsed nitrogen laser operating at 337 nm was used to generate the MALDI ions. All spectra are the result of signal averaging of between 100 and 200 shots. Data processing was done with the IGOR Pro software package (Wavemetrics Inc., Lake Oswego, OR).

The *E. coli* digests were analyzed on a hybrid QSTAR quadrupole/time-of-flight (Qq-TOF) system, equipped with an orthogonal MALDI source (Applied Biosystems/MDS SCIEX, Concord, Canada).

The plasma fraction was analyzed by on-line liquid chromatography/electrospray ionization mass spectrometry (LC/ESI-MS), on a Finnigan LCQ$^{\mathrm{Deca}}$ ion trap mass spectrometer (Finnigan, San Jose, CA). The column was a Micro-Tech Scientific reversed-phase (5 µm) n-octylsilica column (Heseria, CA), with dimensions of 300 µm × 150 mm. A gradient separation was performed using water with 0.5% acetic acid (solvent A) and acetonitrile with 0.5% acetic acid (solvent B). The gradient used was as follows: increase B from 5 to 15% over 3 min, then from 15 to 45% B over 27 min, then increase B to 65% over 10 min. An approximate solvent flow rate of 1 µL/min was obtained by splitting the flow from the HPLC pump.

2.2.7 Safety Considerations

Proper procedures ("Biohazard Level 2") should be followed when handling human plasma samples.

2.3 Results and Discussion

2.3.1 Optimization of Microcolumn Size

From previous work with protein binding and digestion on R2 beads [22], a ratio of 4 μg of beads per picomole of protein was found to produce the best peptide mass maps from the MALDI data. However, for the digestion of proteins in a column, improved results could be obtained if the amount of beads used to pack the column was further lowered. Thus, columns were packed with 30 μg of R2 beads for 100 μL samples consisting of 100-nM protein. For more dilute samples, the column size was reduced to maintain a near-optimal bead to protein ratio. It was found that minimal amounts of packing material produce the best quality peptide mass maps. If the ratio of bead to protein is increased, the quality of the resulting MALDI spectrum is seen to deteriorate. In this case, analysis of the digestion products by MALDI-TOF MS displayed intense trypsin autolysis peaks, which dominate the obtained spectra. It is suspected that the increase in surface area of the stationary phase leaves more exposed sites for trypsin adsorption. Adsorption of the enzyme to the hydrophobic support would likely lead to enzyme denaturation, thereby severely inhibiting enzyme activity. Also, the binding of trypsin to the exposed surface would prevent further complexation with the protein, also preventing digestion. Thus by minimizing the amount of beads used, the hydrophobic surface becomes increasingly saturated with protein sample prior to passing the enzyme though the column. This reduced the extent of enzyme autolysis and improves digestion results.

2.3.2 In-Column Digests of Standard Proteins

Packing the beads in a column offers several advantages over the direct addition of beads to the sample vial. Bead manipulation is not required as the solutions are pumped through the column. Protein reduction and derivatization, as well as sample washing is easily achieved in a column format. The peptides can be eluted from the column into small volumes of solution. In doing so, the MALDI sample preparation can be optimized. By avoiding the direct addition of beads to the MALDI target for analysis, as with previously reported method [22], a lower laser power is used, which helps to increase spectral resolution. Also, an on-probe washing step can be incorporated with this procedure, which further improves spectral quality by removing salts associated with the matrix solution itself.

To illustrate the performance, a 100-μL sample of 100-nM cytochrome c was loaded onto a column and digested for 30 min. The tryptic peptides were then eluted and analyzed by MALDI MS. The obtained MALDI spectrum is displayed in Figure 2.1A. From the spectrum, one can see that many peptide fragments are visible over the mass range of 750-2250 Da. A total of 28 peaks were assigned to fragments of cytochrome c, representing 91% sequence coverage. In other experiments, samples of 100 nM bovine serum albumin (BSA) digested in column generated a peptide map with 43 peptides, giving 69% sequence coverage, while a 100 nM lysozme solution digested in column yielded 28 peptides, giving 100% sequence coverage. These results demonstrate that dilute proteins can be concentrated and digested in-column and that the resulting peptide fragments can be effectively eluted and analyzed by MALDI TOF mass spectrometry.

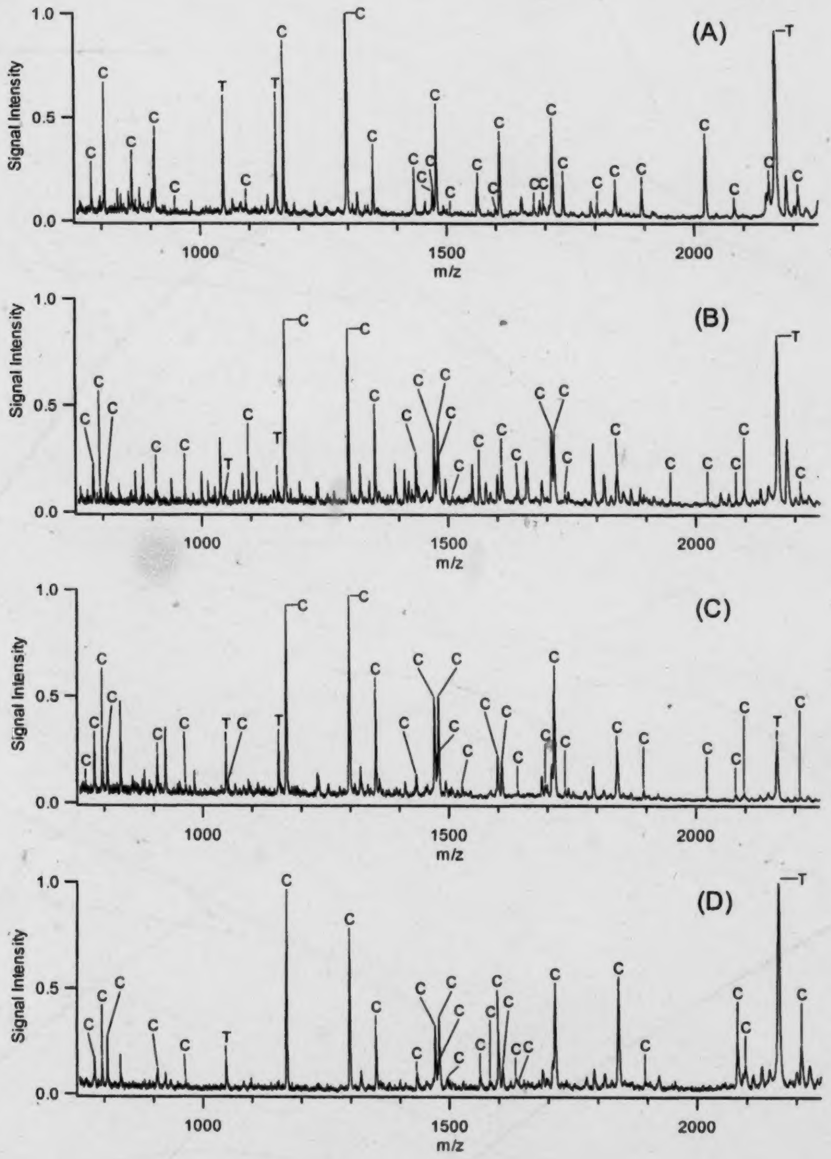


Figure 2.1 MALDI spectra showing the products of the digestion of 100 nM cytochrome c incolumn that was packed with R2 beads for (A) a clean solution, 28 peaks were detected, giving 91% coverage, (B) a solution containing 2M NaCl, 27 peaks were detected, giving 100% coverage, (C) a solution containing 2M urea, 26 peaks were detected, giving 97% coverage, and (D) with 0.01% SDS in solution, 25 peaks were detected, giving 97% coverage. C = peptide fragments of cytochrome c, T = peptide fragments from trypsin autolysis.

Simple protein mixtures can also be digested using the microcolumn digestion procedure. The standard protein α-casein, obtained as a purified product from Sigma, contains both α_{S1}- and α_{S2}-casein. Microcolumn digestion of a 200 nM solution of α-casein revealed peptide fragments from each of these two proteins. For α_{S1}-casein, a total of 33 peaks were detected, giving 100% sequence coverage for the protein. The α_{S2}-casein in the sample resulted in the detection of 20 peptide fragments, giving 63% sequence coverage (data not shown).

2.3.3 Samples Contaminated with NaCl, Urea, and SDS

Inorganic and organic salts such as NaCl and urea are commonly used in protein sample work-up. High salt concentrations are known to interfere with the ionization process and create adducts with the peptides, and therefore sample clean-up prior to MALDI analysis becomes necessary. Solutions of BSA, cytochrome c, or lysozyme, at a concentration of 100 nM, were prepared in 2 M NaCl or 2 M urea. The solutions were flowed through columns packed with R2 beads. Prior to digestion, the column was washed with 100 μL of water to remove any weakly bound contaminants. The peptide mass maps resulting from the digestion of the protein samples were obtained. Figure 2.1B and 2.1C displays the MALDI spectra obtained from the digestion of cytochrome c contaminated with 2M NaCl or with 2M urea, respectively. For the sample containing NaCl, a total of 26 peptides were assigned to fragments of cytochrome c, representing 100% sequence coverage. The urea-containing sample displayed 27 peaks, representing 97% sequence coverage. The quality of the MALDI spectra for the salt- and urea-containing samples is essentially constant when compared to the uncontaminated cytochrome c digest (Figure 2.1A). Such contaminants will largely not be retained on the hydrophobic resin in the column during the sample loading step, and are further removed from the column by incorporation of a washing step prior to digestion and therefore do not deter the MALDI analysis. Results comparable to the uncontaminated samples of BSA and lysozyme were also achieved for these proteins in the presence of 2M NaCl and 2M urea.

Sodium dodecyl sulfate (SDS) is often added to protein samples to aid in the solubilization of the protein. The surfactant interacts with the protein at hydrophobic domains, facilitating solubilization. Due to these interactions, SDS may also partially hinder protein binding to the reversed-phase packing material of the column. Figure 2.1D show the MALDI spectrum of the 100 nM cytochrome c sample prepared with 0.01% (w/v) SDS. As seen from the figure, the signal intensity has decreased as compared to the previous samples. However, 25 peaks were still detected, representing 97% sequence coverage. Although the presence of SDS in the sample has an affect on the quality of the resulting MALDI spectrum, a reliable peptide mass map was still obtained. It is noted here that the presence of 0.01% SDS in the sample represents a modest ratio of SDS to protein, with an 80 fold mass excess of SDS over cytochrome c in the solution. The method can therefore tolerate modest amounts of SDS, relative to the low protein concentration.

2.3.4 Protein Digestion at Shorter Reaction Times

The reaction rate for a particular enzymatic digestion of a protein is dependant on both the protein and enzyme concentration. The in-column digestion procedure incorporates a sample preconcentration step prior to digestion. Therefore, the concentration of the protein (as well as the enzyme) on the surface is high. As a result, longer digestion times are not required in order to achieve high proteolysis. The digestion times were varied in order to optimize the intensity and number of peaks for each spectrum. Figure 2.2 displays the MALDI spectra obtained from the digestion of 100-nM lysozyme at 5 and 30 min. The 30 min digest, shown in Figure 2.2A, revealed 28 peptide fragments of lysozyme, while the 5 min digest produced 25 detectable peptide fragments (Figure 2.2B). The intensities of the spectra were similar and the sequence coverage was 100% for each digest. Although a 30 min incubation time gave a slightly higher number of detectable peaks, these results show that 5 min was sufficient to produce a very complete digestion. Similar results were obtained for cytochrome c and BSA digested at these times. It was concluded that 30-min digestion time was more than sufficient for most in-column digestion applications.

2.3.5 Concentration Limit for In-Column Digestion

To detect and analyze lower concentration of the protein, the size of the column bed needs to be decreased. Figure 2.3 displays the MALDI spectrum from the digestion of 25 nM BSA, digested for 30 min. The column was packed with 10 μg of R2 beads and 310 ng of trypsin was used. A total of 30 peptides were detected resulting in sequence coverage of 45%. This number is somewhat lower than that found from the

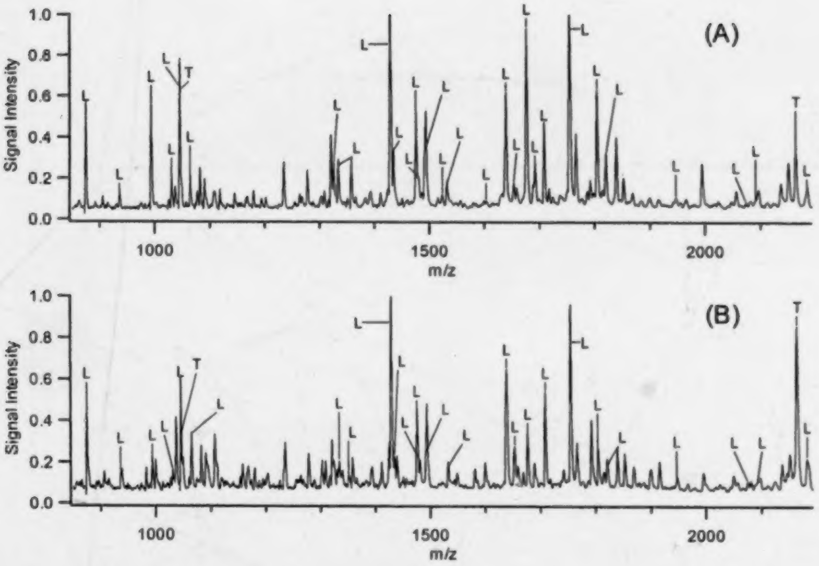


Figure 2.2 MALDI spectra showing the products of digestion of 100 nM lysozyme in-column packed with R2 beads using (A) 30 min digestion time and (B) 5 min digestion time. L = peptide fragments of lysozyme, T = peptide fragments from trypsin autolysis. Insert displays the MALDI spectrum from 2500 Da to 4500 Da.

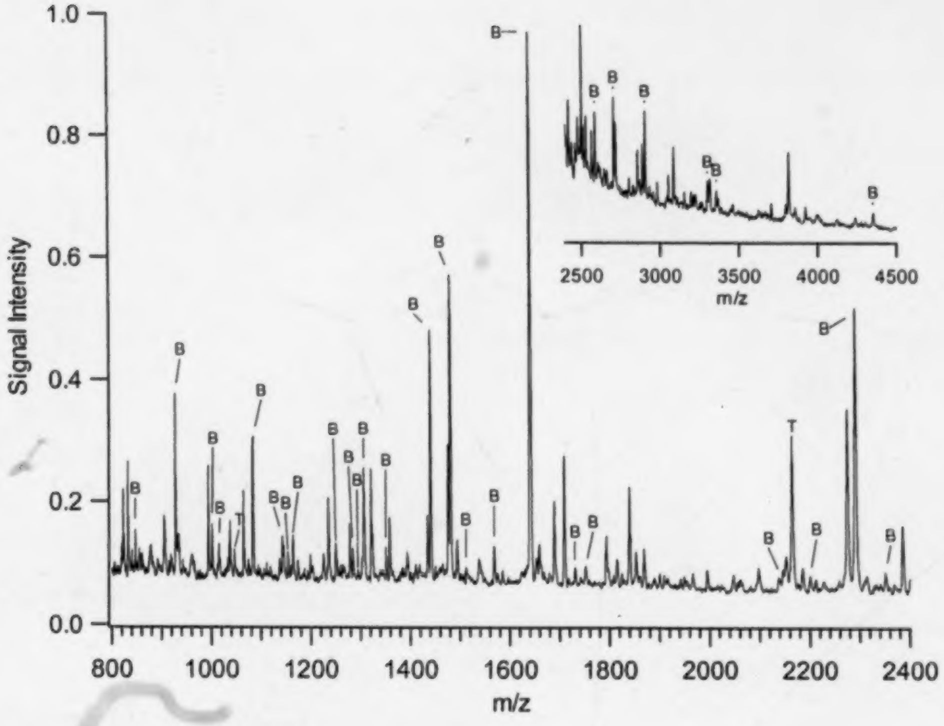


Figure 2.3 MALDI spectrum obtained from the digestion of 25 nM BSA in-column. B = peptide fragments of BSA, T = peptide fragments from trypsin autolysis.

digestion of 100 nM BSA, with 43 peptides, giving 69% sequence coverage. However, a 100 μL solution of 25 nM cytochrome c digested in-column yielded 25 detectable peptide fragments from MALDI analysis of the digestion products, giving 97% sequence coverage. This is essentially identical to that found from the 100 nM cytochrome c digest. The overall intensity and number of peaks from these digests suggest that digestion of a lower concentration is possible. However, as the concentration is further lowered, the possibility of sample contamination becomes an increasing concern. For this reason, peaks corresponding to keratin peptide fragments are often seen with increasing intensity at lower protein concentrations.

2.3.6 Anion Exchange Fractionation of *E. coli*

To identify proteins from a complex protein mixture, a separation of the sample prior to analysis is often required. Ion exchange liquid chromatography is often used to initially fractionate protein mixtures as this method can separate a variety of proteins over a wide range of molecular weight and hydrophobicity. However, direct analysis of ion exchange fractions is hampered by the high concentration of salt that is required to perform the separation. Sample cleanup, as well as preconcentration, must therefore be performed for subsequent sample workup of an ion exchange fraction. The in-column digestion procedure described here provides an ideal method for the direct analysis of protein fractions from ion exchange separations.

To illustrate the application of in-column protein capture and digestion of HPLC fractions, an *E. coli* extract was subjected to anion exchange chromatography using the procedure described in Section 2.2.2. As an example, a 100 μL aliquot of fraction #29 (*i.e.*, the fraction collected at 29-30 min during the separation) was loaded on a microcolumn and subjected to the in-column digestion procedure. This fraction contained approximately 0.3 M NaCl from the elution buffer of the ion exchange separation. Following tryptic digestion of fraction 29 in the microcolumn, the peptides were eluted from the microcolumn and spotted on a MALDI plate for analysis on a hybrid Qq-TOF system. The total MS scan obtained from the MALDI analysis of this sample is displayed in Figure 2.4A. For comparison, an aliquot of fraction 29 was also digested with trypsin using a solution-phased digestion protocol, with 2 hours digestion time, and analyzed on the MALDI Qq-TOF system. A ZipTip (Millipore) cleanup of the digested sample was required to obtain detectable peptide signals, and this spectrum is displayed in Figure 2.4B. MS/MS fragmentation spectra were then collected from several of the peaks detected in the total MS scans. The resulting spectra were used to search the *E. coli* database with MASCOT (www.matrixscience.com). The database search yielding MOWSE scores with significant homology with the matched peptides are summarized in Table 2.1.

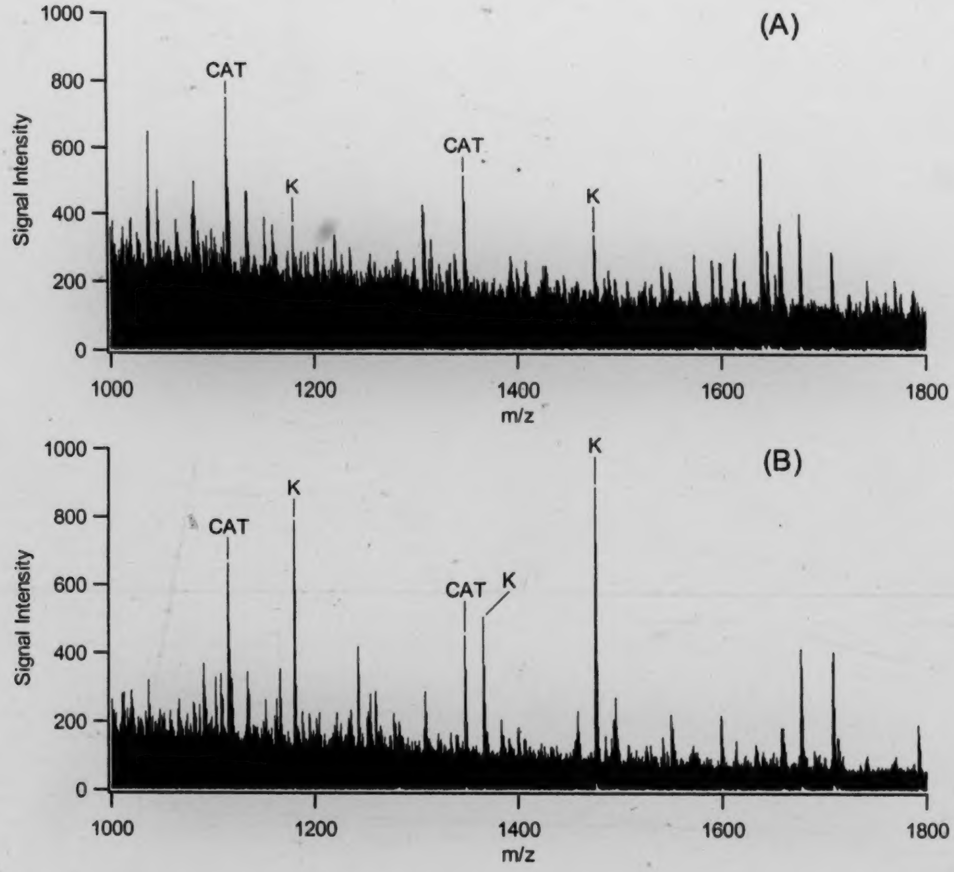


Figure 2.4 Total MS scans from MALDI analysis of the digestion products of an *E. coli* fraction from an ion exchange separation on a hybrid quadrupole-time-of-flight system. (A) is for the *E. coli* sample that was digested in-column and (B) is the same *E. coli* fraction digested using a traditional solution phase protocol, with ZipTip cleanup. CAT = peptide fragments identified by MS/MS fragmentation as *E. coli* chloramphenycol acetyltransferase, K = keratin fragments.

From the results shown in Table 2.1, the column digestion procedure generated two unique peptides, with masses 1114.5 Da and 1347.7 Da, which were tentatively identified by MASCOT. Each of these peptides is from the same protein, namely chloramphenycol acetyltransferase (CAT), SWISS-PROT AC P00483. The MOWSE score from each of these peptides indicated significant homology in each case. The presence of two CAT peptide fragments in the sample gives a confident identification of the protein in fraction 29. These two peptides were also observed in the solution digestion, as shown in Figure 2.4B. The MOWSE scores from MS/MS fragmentation of these peptides were essentially identical to those generated from the column digestion procedure (see Table 2.1).

Three additional peptides were also detected with high intensities in the solution digest that could not be identified by MASCOT, searching the *E. coli* database. A subsequent search of these peptides under the human database subset provided tentative identifications. From Table 2.1, the peptides with masses 1179.7, 1365.7, and 1475.8 were all identified as fragments of keratin. Peaks with masses corresponding to two of these peptides (1179.7 and 1475.8 Da) were also observed in the MALDI spectrum of the

Table 2.1 Summary of the MASCOT Search Results for Peptides Generated from a Digest of an *E. coli* Fraction, Obtained from an Anion Exchange Separation of the Total Cell Extract[a].

	Peak (Da)	Mowse Score	Probability	Identified Protein
Column digest	1114.5	27	> 20 significant homology > 28 extensive homology	Chloramphenycol acetyltransferase
	1347.7	26	> 15 significant homology > 27 extensive homology	Chloramphenycol acetyltransferase
Solution digest with ZipTip	1114.5	29	> 23 significant homology > 31 extensive homology	Chloramphenycol acetyltransferase
	1347.7	27	> 18 significant homology > 29 extensive homology	Chloramphenycol acetyltransferase
	1179.7	44	> 29 significant homology > 31 extensive homology	Keratin Type II
	1365.7	34	> 27 significant homology > 33 extensive homology	Keratin Type I
	1475.8	20	> 31 significant homology > 36 extensive homology	Keratin Type II

[a] The sample was digested using the column digestion procedure, or a traditional solution-phased tryptic digestion, followed by ZipTip cleanup of the sample. The samples were analyzed by MALDI MS on a hybrid Qq-TOF instrument.

sample that was digested in-column, but with significantly lower intensity. The third peptide (1365.7 Da) was completely absent in the column-digestion spectrum.

From this example, it is clear that the in-column digestion procedure is a useful method for the direct identification of proteins from ion exchange separations. The method is comparable to a traditional solution phase digestion, with ZipTip cleanup of the sample. However, the column digestion procedure requires a shorter digestion time and usually results in a lower abundance of detectable keratin fragments. Moreover, column digestion has the potential to be fully integrated with ion exchange column separation of proteins for on-line protein identification.

2.3.7 Analysis of Human Plasma

To demonstrate the versatility of this method for analyzing complex protein mixtures, a fraction of human plasma was subjected to the in-column digestion procedure. The plasma sample was fractionated by the Cohn precipitation method to obtain a sample that was not overwhelmed by the abundance of human serum albumin (HSA), which is present in plasma at an extremely high relative concentration, compared to the minor protein components of the plasma. Cohn fraction IV-4 was therefore used, as this fraction still contains an abundance of proteins, in the presence of a reduced amount of HSA. The precipitate was re-dissolved in 1 mL of 50 mM NH_4HCO_3 (initial volume of plasma was 1 mL), and then further diluted in 50 mM NH_4HCO_3 by a factor of 1:100. The diluted fraction was then reduced and alkylated prior to loading onto the column. A total of 100 μL of the diluted plasma fraction was loaded on the microcolumn, followed by 100 μL of water to wash the column prior to digestion. A buffered trypsin solution was then added and the column was allowed to incubate for 30 min. Following the digestion, the peptides were eluted from the column with 5 μL of 50% acetonitrile/ water with 0.1% TFA (v/v), and the extract was partially dried down in a SpeedVac to reduce the acetonitrile content of the sample. The partially dried extract was diluted in 5 μL of water, and the entire sample was then injected into the LCQ^{Deca} for analysis.

Figure 2.5 displays the base peak chromatogram from the LC/ESI analysis of the digested human plasma fraction. Representative MS/MS spectra from the base peak chromatogram, along with the tentative identification of these spectra, are also shown in the figure. MS/MS data files were searched using the SEQUEST algorithm [30], under the human database subset. The output of the SEQUEST search was evaluated according to the criteria described by Yates [31] in order to filter out the results corresponding to valid identifications. The plasma proteins identified by SEQUEST are summarized in Table 2.2.

From Table 2.2, a total of 14 unique polypeptide chains were identified from the plasma protein fraction. This list includes various immunoglobulin heavy and light chains, apolipoproteins, transferrin, serum albumin, and other plasma proteins. Keratin was also observed, likely as a contaminant during blood collection or later in sample handling. Of these polypeptide chains, many were identified from at least 2 peptides, thus significantly increasing the confidence of the identification. For example, from the SEQUEST search results, a total of 17 peptides fragments were assigned to apolipoprotein A-I, which yields sequence coverage of 70%. From Table 2.2, three

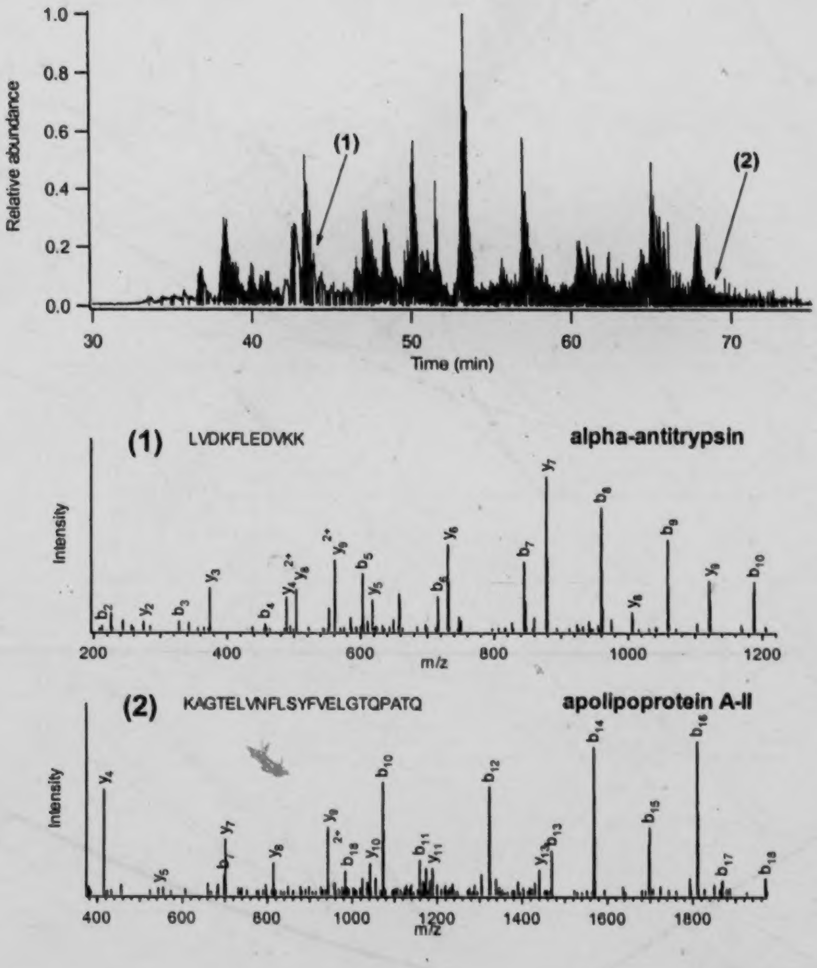


Figure 2.5 Base peak chromatogram obtained from the analysis of a diluted plasma fraction following digestion by LC/ESI. Two representative MS/MS spectra are also shown, along with the tentative sequence of the fragmented peptide, and identity of the parent protein. The plasma sample was digested with trypsin using the in-column digestion procedure.

Table 2.2 Summary of the polypeptides identified using MS/MS data from a diluted human plasma fraction, as obtained from the SEQUEST search algorithm [a].

Polypeptides Identified	Number of Peptides Observed
apolipoprotein A-1	17
serum albumin	14
α1-antitrypsin	12
transferrin	9
immunoglobulin χ heavy chain	5
keratin[b]	5
immunoglobulin λ light chain	4
immunoglobulin α_1 heavy chain	4
haptoglobulin I	3
immunoglobulin κ light chain	2
apolipoprotein A-II	2
immunoglobulin α_2 heavy chain	1
apolipoprotein A-II	1
α_2-macroglobulin	1

[a] The number of peptides found corresponding to each polypeptide is indicated in the table. The samples were analyzed by LC/ESI MS on an LCQ ion trap instrument.

[b] Keratin was observed as a contaminant during blood collection or from sample handling.

polypeptides were identified from only a single peptide that fit the criteria used to filter the output from the SEQUEST database search. However the extent to which the MS/MS spectra matches the predicted peptide sequence provides for confident identification of these proteins in the fraction.

The concentration of proteins present in this mixture were not precisely known, however a maximal value can be given based on the normal levels of these proteins in plasma [31]. Of the identified proteins, most are present in this fraction at sub-micromolar levels. Serum albumin makes up the only possible exception, where the maximum concentration of this protein in Fraction IV-4, assuming 100% recovery, is 6 μM. Other proteins are present at significantly lower concentrations. The protein α_2-macroglobulin has a maximum predicted concentration of 33 nM. The actual concentrations of all components in this fraction are expected to be even lower than predicted, since the recovery of these proteins in the precipitated fraction is not quantitative. Even at these low concentrations, the column digestion procedure generated an abundance of detectable peptide fragments, as observed by the base peak chromatogram of Figure 2.5.

It is worth noting that the protein contained in the 100 μL diluted plasma fraction is expected to exceed the capacity of the column. Based on the binding capacity of Poros R2 beads, a microcolumn packed with 30 μg of stationary phase has a capacity of ~20 pmol (for a 50 kDa protein). In other terms, a 100 μL protein solution at a concentration of 200 nM is sufficient to saturate the column. Although the column will not capture all the protein from the plasma fraction, the amount of protein that does adsorb to the hydrophobic support of the microcolumn is sufficient to generate several peptide fragments that yield high quality MS/MS spectra.

The results shown above demonstrate that the in-column digestion procedure can be used for sample cleanup, concentration, and digestion of a relatively complex biological sample mixture containing proteins present at sub-micromolar concentrations.

2.3.8 Analysis of Human Tears

Human tears were collected in house. The sample was reduced and alkylated at pH 8 prior to loading into the microcolumn. The pH adjustment precipitated proteins out of solution. The sample was centrifuged and the supernatant was used for analysis. The supernatant was diluted 10 times and 100 μL was loaded onto the microcolumn. Following digestion the eluent was injected in a LCQ^{Deca} system for analysis. The base peak chromatogram is displayed in Figure 2.6. Tear pre-albumin, lactoferrin, and IgA were identified by the SEQUEST searching algorithm. Table 2.3 provides a summary of the proteins identified and displays the relative concentrations of abundant proteins in human tears. Three of the four abundant proteins present in tears were detected by this method. Lysozyme was not seen, presumably because this protein would precipitate at pH 8 at high concentrations.

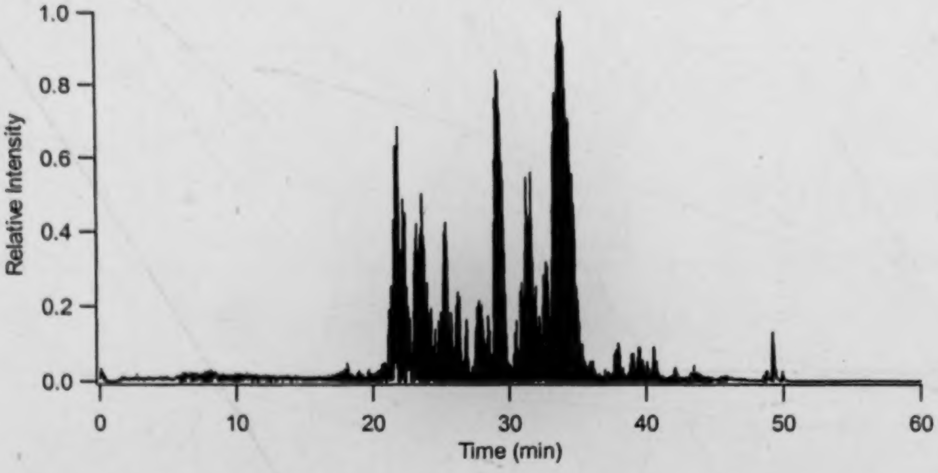


Figure 2.6 Base peak chromatogram obtained from the analysis of a diluted tears from human following in-column digestion by LC/ESI.

Table 2.3. List of the four most abundant proteins in tears.

Tear Proteins	Concentration of 100 μL aliquot (nM)	# Peptides identified
Tear pre-albumin	325	2
Lysozyme	85	0
Lactoferrin	25	1
IgA	10	1

2.4 Conclusions

The method of preconcentration and digestion of proteins while adsorbed on hydrophobic microbeads in a column produces many peptide fragments that can be detected by MALDI or ESI mass spectrometry. High sequence coverage was obtained from MALDI peptide mass maps starting with as low as 25 nM protein solutions. The method is resistant to high concentrations of salt and urea, and moderate relative concentrations of SDS to protein. The increase in protein concentration obtained by adsorbing the sample onto a surface prior to digestion improves reaction kinetics, thus allowing for the digestion time to be reduced to less than 30 min. The method is ideally suited for direct analysis of protein fractions separated by ion exchange, providing results for an *E. coli* extract that were comparable to a traditional solution digestion followed by Ziptip cleanup. The method was also applied to the analysis of a human plasma fraction by LC/ESI, where 14 unique polypeptides were identified from analysis of a single column digest. Human tears were also used as an application for the on-column digestion method. The resulting digest identified three of the four abundant proteins present in human tears. It is clear that the in-column protein capture and digestion procedure can be used as a means of handling dilute protein solutions containing high amounts of salts and other impurities. The entire procedure can potentially be automated for parallel processing of a large number of proteins in proteomics research. The procedure can also be integrated into an on-line, fully automated system where a protein sample is first separated by chromatography, followed by microcolumn protein capture/preconcentration/washing and subsequent digestion. The resulting digests can be analyzed by LC MS/MS.

2.5 Literature Cited

[1] Mann, M; Hendrickson, R. C.; Pandey, A. *Annu. Rev. Biochem.* **2001**, *70*, 437-473.

[2] Aebersold, R.; Goodlet, D. R. *Chem. Rev.* **2001**, *101*, 269-295.

[3] Gevaert, K.; Vandekerckhove, J. *Electrophoresis* **2000**, *21*, 1145-1154.

[4] Larsen, M. R.; Roepsorff, P. *Fres. J. Anal. Chem.* **2000**, *366*, 677-690.

[5] Chaurand, P.; Luetzenkirchen, F.; Spengler, B. *J. Amer. Soc. Mass. Spectrom.* **1999**, *10*, 91-103.

[6] Kellner, R. *Fres. J. Anal. Chem.* **2000**, *366*, 517-524.

[7] Figeys, D.; Pinto, D. *Electrophoresis* **2001**, *22*, 208-216.

[8] Quadroni, M.; James, P. *Electrophoresis* **1999**, *20*, 664-677.

[9] Shevchenko, A.; Wilm, M.; Vorm, O.; Mann, M. *Anal. Chem.* **1996**, *68*, 850-858.

[10] Li, J.; Wang, C.; Kelly, J. F.; Harrison, D. J.; Thibault, P. *Electrophoresis* **2000**, *21*, 198-210.

[11] Lazar, I.; Ramsey, R.; Ramsay, J. M. *Anal. Chem.* **2001**, *73*, 1733-1739.

[12] Ericsson, D.; Ekstrom, S.; Nilsson, J.; Bergquist, J.; Marko-Varga, G.; Laurell, *Proteomics* **2001**, *1*, 1072-1081.

[13] Litborn, E.; Emmer, A.; Roeraade, J. *Anal. Chimica Acta* **1999**, *401*, 11-19.

[14] Litborn, E.; Emmer, A; Roeraade, J. *Electrophoresis* **2000**, *21*, 91-99.

[15] Whittal, R. M.; Keller, B. O.; Li, L. *Anal. Chem.* **1998**, *70*, 5344-5347.

[16] Amankwa, L. N.; Kuhr, W. G. *Anal. Chem.* **1992**, *64*, 1610-1613.

[17] Licklider, L.; Kuhr, W. G. *Anal. Chem.* **1994**, *66*, 4400-4407.

[18] Blackburn, R. K.; Anderegg, R. J. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 483-494.

[19] Licklider, L.; Kuhr, W. G. *Anal. Chem.* **1998**, *70*, 1902-1908.

[20] Marie, G.; Serani, L.; Laprevote, O. *Anal. Chem.* **2000**, *72*, 5423-5430.

[21] Vecchione, G.; Casetta, B.; Santacroce, R.; Margaglione, M. *Rapid Comm. Mass Spectrom.* **2001**, *15*, 1383-1390.

[22] Doucette, A.; Craft, D.; Li, L. *Anal. Chem.* **2000**, *72*, 3355-3362.

[23] Cohn, E. J.; Strong, L. E.; Hughes Jr., W. L.; Mulford, D. J.; Ashworth, J. N.; Melin, M.; Taylor, H. L. *J. Am. Chem. Soc.* **1946**, *68*, 459-475.

[24] Li, M. X.; Saude, E. J.; Wang, X.; Pearlstone, J. R.; Smillie, L. B.; Sykes, B. D. *Eur. Biophys. J.* **2002**, *31*, 245-256.

[25] Wang, Z.; Russon, L. M.; Li, L.; Roser, D. C.; Long, S. R. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 456-464.

[26] Gobom, J.; Nordhoff, E.; Ekman, R.; Roepstorff, P. *Int. J. Mass Spectrom. Ion. Processes* **1997**, *169/170*, 153-163.

[27] Annan, R. S.; McNulty, D. E.; Carr, S. A. In *Proc. 44th ASMS Conf. on Mass Spectrom. Allied Topics*, Portland, OR, **1996**, 702.

[28] Craft, D.; Doucette, A.; Li, L. In *Proc. 48th ASMS Conf. on Mass Spectrom. Allied Topics*, Long Beach, CA, **2000**, 1429.

[29] Whittal, R. M.; Li, L. *Anal. Chem.* **1995**, *67*, 1950-1954.

[30] Eng, J.; McCormack, A. L.; Yates III, J. R. *J. Am. Mass Spectrom.* **1994**, *5*, 976-989.

[31] Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates III, J. R. *Nat. Biotechnol.* **1999**, *17*, 676-682.

Chapter 3

Mass Spectrometric Study of the Effects of Hydrophobic Surface Chemistry and Morphology on the Digestion of Surface-Bound Proteins [3,4]

3.1 Introduction

There has been a long-standing interest in performing enzymatic digestion of proteins while the sample is bound to a surface or substrate for the purpose of protein identification by either Edman sequencing or by mass spectrometry (MS). For example, protease digestion of proteins electroblotted onto nitrocellulose was demonstrated by Aebersold *et al.* in 1987 [1]. In a more recent application, endoprotease digestion of

---

[3] A portion of this chapter has been published as: A Doucette, D Craft and Liang Li "Mass Spectrometric Study of the Effects of Hydrophobic Surface Chemistry and Morphology on the Digestion of Surface-Bound Proteins" *J. Am. Soc. Mass Spectrom.*, **2003**, *1*, 203-214.

[4] The bulk of the results described in this chapter were conducted with equal contributions from A. Doucette and D. Craft. Specifically, these include for all data related to the investigation of different reversed-phase media for protein adsorption and digestion.

surface-bound proteins has been applied to the technique of surface-enhanced laser desorption ionization (SELDI) MS [2-4]. The SELDI technique involves the use of surfaces with various functional activities to capture target proteins. Following protein capture, endoprotease digestion can be directly performed *in situ*, with the resulting peptides analyzed by MS.

In previous work, I have demonstrated the utility of a protein-digestion protocol involving an initial capture of proteins onto reversed-phase (hydrophobic) chromatography beads, followed by tryptic digestion of the bound protein for MS analysis [5-7]. The research goal is to develop an integrated system for automated protein sample cleanup, preconcentration, digestion, and subsequent MS analysis of the digestion products for protein identification. Although earlier results reveal that surface-bound proteins will undergo enzymatic cleavage, the specific effects of protein adsorption, surface chemistry, and morphology on the digestion process have not yet been explored.

Considering that protein sequence information obtained by MS is strongly dependent on the extent of protein digestion, understanding the digestion process of surface-bound proteins is clearly important in developing an effective protein identification and characterization method involving surface digestion. However, to my knowledge, there has been no systematic investigation on the effects of varying hydrophobic surface chemistry and morphology on the digestion of surface-bound proteins. Although a number of reports have demonstrated the feasibility of digesting proteins while bound to various types of membrane surfaces [1,8-17], there are conflicting results on the extent of proteolysis that has been observed on membranes. For example, Scleuder *et al.* reported on their digestion results using hydrophobic versus hydrophilic membranes in the form of modified or unmodified PVDF, concluding that superior peptide mass fingerprints can be obtained using a hydrophilic membrane [14]. It was shown that the digestion results on hydrophobic membranes were poor, providing irreproducibly positive results at best. In another report, however, Bienvenut *et al.* observed that *in situ* digestion on a hydrophobic membrane generates similar peptide mass maps as those obtained from in-gel digestion [9]. These conflicting results may well reflect the differences in experimental conditions used for digestion of membrane-bound proteins.

For bead-digestion application [5-7], the experimental conditions used are quite different from those described in the above-mentioned reports. An intuitive consequence of protein adsorption to a hydrophobic bead surface is that interactions between the protein and surface may lead to inaccessibility of the enzyme to cleave certain digestion sites of the protein. Supporting evidence for this was given by Aguilar *et al.*, who found that for a protein adsorbed to reversed-phase bonded silica, only those sites exposed to the solvent were accessible to proteolysis [18]. They reported that a tryptic digestion of cytochrome c adsorbed to reversed-phase C_{18} beads was blocked at certain amino acid residues that correspond to regions of protein incontact with the surface. In some related work, proteins bound to antibodies have been digested with enzymes as a means of mapping epitopes [19-21]. Chemical modification of free *vs.* antibody-bound proteins has also been used to study protein-binding sites [22]. It has been shown that the rate of enzymatic cleavage or chemical reactivity for a protein complex is reduced, most significantly at regions where the protein is in contact with the antibody (the epitope).

Similarly previous results from our group on biotinated proteins complexed to avidin-coated beads revealed that such interactions prevented efficient trypsin digestion [23,24]. However, these interactions are quite different from those that are experienced by a protein adsorbed to a hydrophobic surface.

There were also reports indicating that the interactions between a protein and a surface may improve the overall digestion efficiency. Litborn *et al.* observed an increase in the digestion efficiency for protein samples contained in nanovials and hypothesized that the high surface-to-volume ratio of the nanovials helped to promote surface-induced protein denaturation [25]. In Aguilar's report on the digestion of proteins on a C_{18} surface, it was also suggested that conformational changes could result in certain residues becoming increasingly accessible, thereby increasing digestion efficiency at these sites [18]. Guillochon *et al.* made use of a surface digestion protocol to generate unique peptide fragments of hemoglobin through chymotryptic and V8 protease hydrolysis of the protein adsorbed on an anionic resin [26,27].

In this chapter, I present my studies on the digestion characteristics of surface-bound proteins. In particular, the digestion of a protein on different types of hydrophobic supports is explored. Optimal conditions for digestion of adsorbed proteins on these surfaces are therefore determined. In addition, it is demonstrated that digestion on a reverse phase media can be achieved with different enzymes. Using a combination of direct matrix-assisted laser desorption ionization (MALDI), liquid chromatography (LC)/offline MALDI, and LC/electrospray ionization (ESI) MS analysis, subtle changes in protein digestion characteristics are revealed.

3.2 Experimental

3.2.1 Materials and Reagents

The proteins used in this study, including the enzyme trypsin (from bovine pancreas, TPCK treated to reduce chymotrypsin activity; dialyzed, lyophilized) were from Sigma Aldrich Canada (Ontario, Canada). Poros 20 R2 beads (20 μm) were from PerSeptive Biosystems (a gift of Professor F. Cantwell, University of Alberta). PRP-3 (12–20 μm), PRP-1 (12–20 μm), and PRP-Infinity (4 μm) beads were purchased from Hamilton Company (Reno, NV, USA). Reversed-phase C_{18}, C_8, and C_4 bonded-phase silica beads (20-30 μm) were from Vydac (Hesperia, CA, USA). The 1-200 μL microcapillary tips (gel loader tips) were purchased from Rose Scientific (Alberta, Canada). The glass micro-fiber filters were from Whatman (Maidstone, UK). A BioRad protein assay kit based on the Bradford Protein Assay was used to quantify the amount of protein in solution. Analytical grade acetone, methanol, acetonitrile, acetic acid, and trifluoroacetic acid (TFA) were from Caledon Laboratories (Alberta, Canada). Water used in the experiments was from a NANOpure water system (Barnstead/Thermolyne). All other chemicals were purchased from Sigma Aldrich Canada.

3.2.2 Digestion of Cytochrome c Adsorbed on Micro-Beads

Poros R2 beads or Vydac C_{18} silica beads were saturated with cytochrome c and subjected to tryptic digestion using a procedure similar to that described by Aguilar *et al.* [18]. The beads (7.5 mg) were first washed in methanol and then in 35% acetonitrile/water (v/v) and gently shaken for 2 hours in a 500 μL solution of 7.5-mg/mL cytochrome c. Following this, the beads were washed with four changes of enzyme buffer (50 mM NH_4HCO_3 containing 2 mM $CaCl_2$). The beads were incubated with shaking in 0.5 mL of enzyme buffer containing 75 μg of trypsin for 20 hours at 37°C. The reaction was terminated with 2 drops of 2 M HCl and the supernatant was retained. The beads were then eluted with 250 μL each of 25%, 50%, and 75% acetonitrile/water (v/v) containing 0.1% TFA, pooling the extracts prior to analysis. 1-μL portions of the supernatant and of the extract were reserved for MALDI analysis. The supernatant and extracts were combined, and concentrated by solvent evaporation with a Savant SpeedVac Concentrator (Fisher Scientific, Ontario, Canada) to a final volume of approximately 200 μL. The sample was then subjected to HPLC fractionation and analysis.

3.2.3 Solution Digestion of Protein

A 10 μL-protein sample, at a concentration of 1 μg/μL, was buffered with 1 μL of 1 M NH_4HCO_3 and digested using 1 μL of 1-μg/μL trypsin. The solution was incubated at 37°C for 2 hours. The digestion was stopped by acidifying the solution with 10% TFA.

3.2.4 Protein Digestion in Microcolumns

The design of the microcolumn is based on work presented by Roepstorff *et al.* [28] and by Annan *et al.* [29]. Briefly, the column is formed from a microcapillary gel loader tip (1-200 μL), which had been pinched at its end with pliers. A piece of glass micro-fiber membrane formed a frit in the microcolumns. The columns were packed with micro-bead resin by transferring an aliquot of the resin, suspended in methanol at a concentration of 10 mg/mL, to the gel loader tip. The procedure for sample loading and digestion is based on my previous work [6,7]. Pressurized nitrogen gas was used to pump the protein samples through the columns at an approximate flow rate of 10 μL/min. Each 100 μL sample was pumped through the column a total of three times. The column was then washed with 100 μL of water. The BSA and lysozyme samples were treated with DL-dithiothreitol (DTT) and iodoacetamide to break disulfide bonds prior to loading on the column, using standard protocols. The absorbed protein was digested by flowing a 10-μL solution of 31-ng/μL trypsin in 50 mM NH_4HCO_3 through the column. A small volume of enzyme solution was left in the column to prevent the beads from drying. The column was incubated for 30 minutes at 37°C, after which the column was eluted with 3 μL of 50% acetonitrile/water (v/v) with 0.1% TFA. Each column was used once and then discarded.

3.2.5 HPLC Separations and Fractionation

Peptides were separated by HPLC on an Agilent Series 1100 chromatographic system (Agilent Technologies, Palo Alto, CA, USA) using a Vydac n-octylsilica (C_8) column (Heseria, CA, USA) with dimensions of 250 × 4.6-mm. Gradient elution was performed using 0.02% (v/v) TFA in water (buffer A) and 0.02% (v/v) TFA in acetonitrile (buffer B). For online mass spectrometric detection, 0.5% acetic acid was used in place of TFA. The gradient used was as follows: hold B at 2% for 5 minutes, then increase from 2% to 50% B over 48 minutes. The flow rate was set at 1 mL/min. For separation of the adsorbed cytochrome c digests, approximately 25% of the total sample was injected on the column. For the solution phase digest, 10 μL of the sample, equivalent to 8.3 μg of cytochrome c, was injected on the column. UV chromatograms were recorded at 210 nm. For offline analysis of the chromatographic separation by MALDI, the effluent was collected in vials at one-minute intervals with an automated fraction collector. The 1 mL fractions were concentrated to approximately 50 μL by solvent evaporation. The concentrated samples were then analyzed by MALDI-TOF MS.

3.2.6 Mass Spectrometry

For MALDI analysis, the matrix α-cyano-4-hydroxycinnamic acid (HCCA) was used. Prior to use, the HCCA was purified by recrystallization from ethanol. A two-layer deposition method, previously reported by my group, was used to prepare the samples [30]. On-probe washing was performed by placing a 1 μL droplet of water on the dried sample spot, and blowing off the water with air after a few seconds. The

MALDI spectra for the in-column digests were collected on a linear time-lag focusing MALDI time-of-flight mass spectrometer, which was constructed at the University of Alberta and has been described in detail elsewhere [31]. The fractions from the HPLC separation were recorded using a Bruker Reflex III MALDI time-of-flight system (Bremen/Leipzig, Germany). All spectra are the results of signal averaging of between 100 and 200 shots. Data processing was performed with the IGOR Pro software package (Wavemetrics Inc., Lake Oswego, OR, USA).

On-line liquid chromatography/electrospray ionization mass spectrometry (LC/ESI-MS) analysis was performed on a Finnigan LCQ^{Deca} ion trap mass spectrometer (Finnigan, San Jose, CA, USA). The effluent from the column was split, allowing approximately 20 μL/min to flow to the ion trap for MS analysis.

3.3 Results and Discussion

3.3.1 HPLC Separation with UV detection of Peptide Fragments

Cytochrome c was chosen as a model protein to study how the digestion efficiency is affected by adsorption of the protein to a hydrophobic support. This small, hydrophilic protein is readily digested by the enzyme trypsin, producing peptides that are easily separated by HPLC as well as detected by mass spectrometry. Cytochrome c was subjected to tryptic digestion following adsorption to R2 or C_{18} beads using the procedure described in Section 3.2.4. For comparative purposes, the protein was also digested with trypsin using a traditional solution-phase digestion procedure. The resulting digestion products were separated by HPLC and the corresponding UV chromatograms are displayed in Figure 3.1.

Figure 3.1A displays the UV chromatogram for the peptide fragments of the solution digest of cytochrome c. The UV chromatograms for the surface-bound digestions on C_{18} and R2 beads are presented in Figure 3.1B and 3.1C, respectively. When compared to the UV chromatogram from the solution digest, it is apparent that the adsorption of cytochrome c to these surfaces alters the digestion process. In particular, the solution digest reveals various peaks early in the separation having strong signal intensities that display only low intensities in the surface digests. Also, the two surface-digests (Figure 3.1B and 3.1C) reveal intense peaks between 35 and 40 minutes, which are essentially absent in the chromatogram of the solution digest (Figure 3.1A). These results lead to the conclusion that the tryptic digestion of cytochrome c absorbed to a surface produces an increased proportion of larger, more incompletely digested peptide fragments, as compared to a solution digestion of the same protein. Increasing digestion times under the experimental conditions used for surface digestion as well as solution digestion did not result in any more complete digestion. This was evident in a series of experiments where digestions were carried out in varying time periods (from 10 min to 24 hours) and the resulting digests were analyzed by both UV chromatography and MALDI MS. No apparent changes in UV chromatograms and MALDI spectra were observed (data not shown). Also, these differences cannot be attributed to individual peptides being irreversibly bound, or lost from the support. In a subsequent experiment, it was found that when the digestion products of a solution digest of cytochrome c were

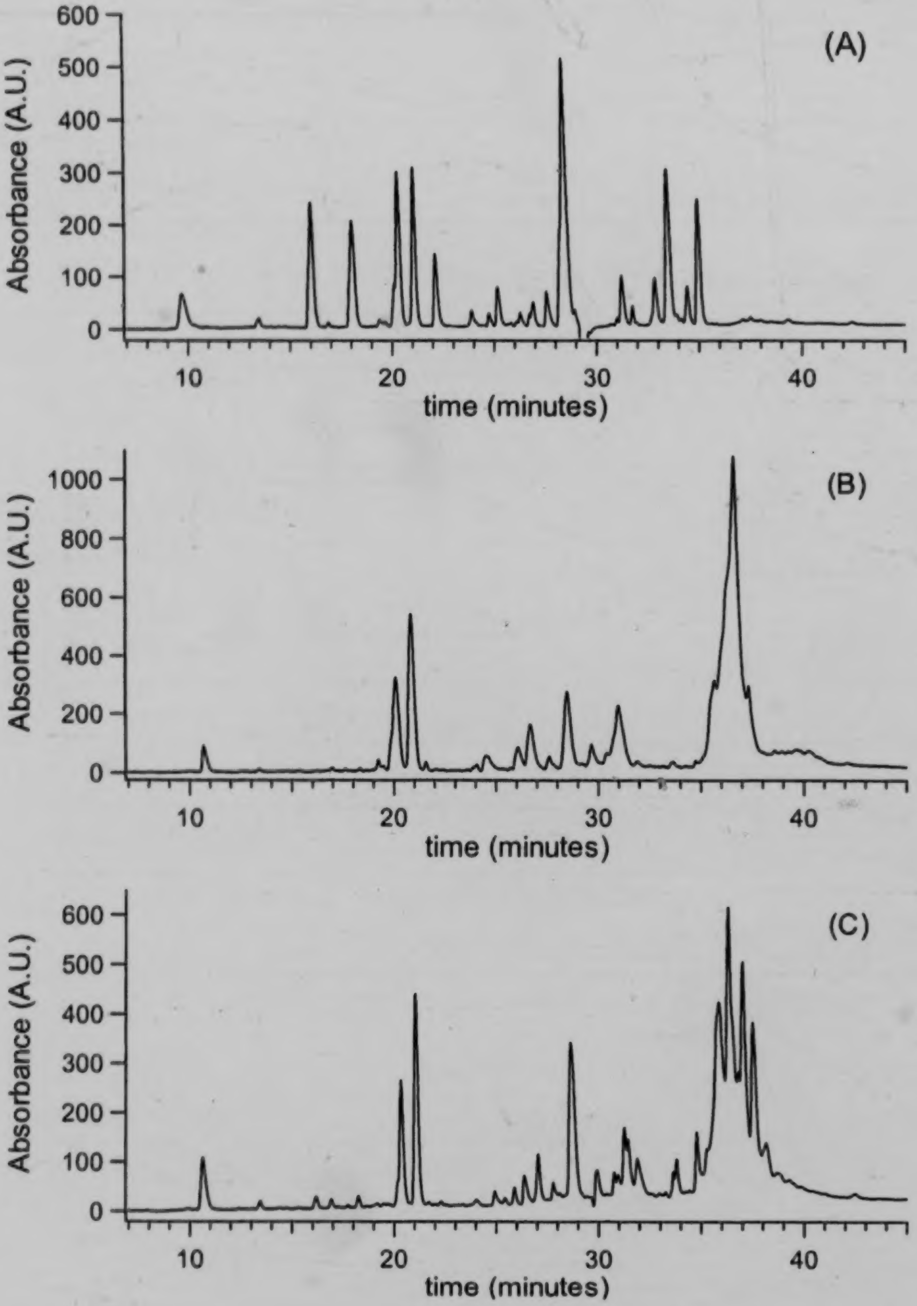


Figure 3.1 UV chromatograms for the HPLC separations of peptide fragments from the digestion of cytochrome c: (A) in solution, (B) on C_{18} silica beads, and (C) on polymeric R2 beads.

loaded and subsequently eluted from the bead surface, no apparent changes were observed in the UV trace compared to the one originally obtained (Figure 3.1A).

A comparison of the chromatograms from the surface-bound digests of cytochrome c on C_{18} and on R2 beads (Figure 3.1B and 3.1C) reveals a closer resemblance. As compared to the UV trace of the solution digest, these two traces display many peaks at similar retention times having similar relative peak intensities. This therefore demonstrates that the relative abundances of many of the peptide fragments in each digest are somewhat similar. Some notable differences are observed between the UV chromatograms from Fig 3.1B and 3.1C. For example, Figure 3.1C (R2 beads) reveals a slightly higher relative amount of early eluting peptides than that of Figure 3.1B (C_{18} beads). Also, the relative intensities of the partly resolved peaks between 35 and 40 minutes in the UV trace are noticeably different in these two traces. These results therefore indicate that for these two different supports, the type of surface has only a minor influence on the digestion process for surface bound proteins.

It should be noted that, in conducting a comparison of digests of a protein on various hydrophobic surfaces, the conditions for digestion were optimized according to each surface used. As mentioned in my previous work [6], improved digestion can be achieved as the bead surface becomes increasingly saturated with protein. The initial experiments involving the comparison of R2 and C_{18} surfaces by HPLC were therefore conducted under completely saturated digestion condition, in which the maximal amount of protein was bound to the bead surface prior to digestion. Using less than saturated conditions will create an additional variable in comparing the effect of surface property on digestion. For example, a non-saturated surface can adsorb and/or denature trypsin, which leads to a less than optimal digestion process, such as impeded enzyme activity and increased trypsin autolysis [6]. Since the surface area and protein adsorptivity are different for different beads (see below), a fixed amount of protein cannot be applied for adsorption onto all beads, as this would result in different protein coverage on different surfaces. Thus, experiments conducted under saturated surface coverage conditions provide the best assessment on the influence of the surface characteristics on trypsin digestion.

To gauge the amount of protein adsorbed on each type of bead surface, the cytochrome c was loaded on 7.5 mg of beads, as described in the Section 3.2.3. Subsequently, the undigested protein was eluted from the bead surface with 75% acetonitrile/0.1% TFA (v/v), the acetonitrile was evaporated by SpeedVac, and the protein was quantified using the Bradford Protein Assay. It was found that, under the conditions used in this work, the amount of cytochrome c adsorbed to the R2 surface was 1.1 mg, and 1.9 mg for the C_{18} surface. It was also found that 1.5 mg, and 2.3 mg of cytochrome c adsorbed to C_4 and C_8 silica beads, respectively. It is clear that different amounts of proteins are adsorbed on different beads. However, these protein amounts represent the saturation of each bead surface. Also, MALDI analysis of the digestion products reveal very weak cytochrome c peak, indicating a small amount of intact cytochrome c still remains after digestion. Since MALDI is not quantitative, the exact amount of the intact protein cannot be determined. No attempt was made to quantify the intact protein by ESI MS, since ion suppression of protein signals in the presence of large amount of peptides are severe and reversed-phase LC separation of proteins from digested peptides are difficult. The presence of undigested cytochrome c in the beads may be attributed to those adsorbed on certain areas of beads that are not accessible by trypsin. Nevertheless, under the saturated digestion conditions, the UV traces can be readily reproduced and do not change significantly as the digestion time varies from 10 min to 24 hours.

While the UV chromatograms shown in Figure 3.1 provide some quantitative comparison of the digestion products and illustrate that the overall digestion efficiency on a C_{18} or an R2 surface, although comparable between each surface, is clearly different than a solution digestion of the same protein. These results do not provide information on the identities of these digestion products. In the following sections, I present the results from MS analysis of the digestion products to compare the number and type of peptide fragments generated by the tryptic digests.

3.3.2 MS Detection of Solution and Surface-Bound Protein Digests

Because of the need to survey as many digested peptides as possible to provide a better amino acid sequence coverage of the protein, three modes of MS detection were carried out for analyzing the products of protein digests, namely LC/ESI, LC/offline MALDI, and direct MALDI, as described in the Section 3.2.6. Peptide fragments of cytochrome c observed under these three detection methods were recorded and the results are summarized in Figures 3.2-3.4. Figure 3.2 displays the peptides observed from the solution digest of cytochrome c, while Figures 3.3 and 3.4 list the peptide fragments detected from the C_{18} and R2 bead digestions, respectively. The first column in these figures lists the predicted cleavage sites from tryptic digestion of cytochrome c.

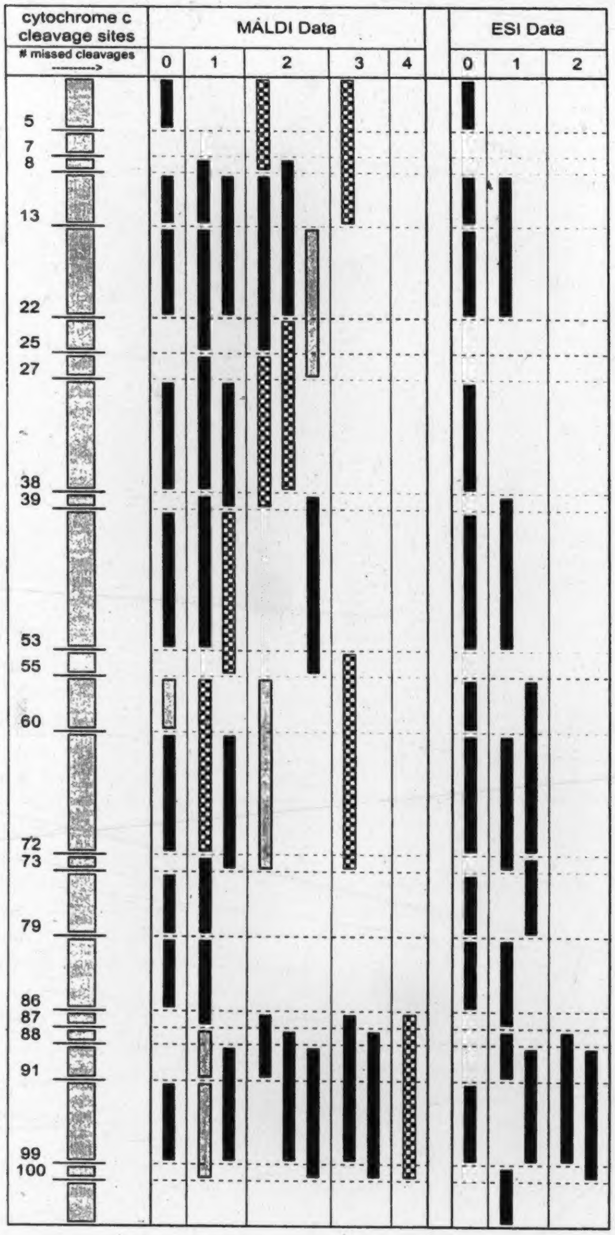

cytochrome c cleavage sites
# missed cleavages
MALDI Data
ESI Data
0
1
2
3
4
0
1
2
5
7
8
13
22
25
27
38
39
53
55
60
72
73
79
86
87
88
91
99
100

Figure 3.2 Tryptic peptides detected from the digest of cytochrome *c* in solution by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a gray bar, while those exclusive to LC/off-line MALDI are indicated with a checkered bar. Fragments commonly observed in each MALDI analysis method is indicated with a black bar.

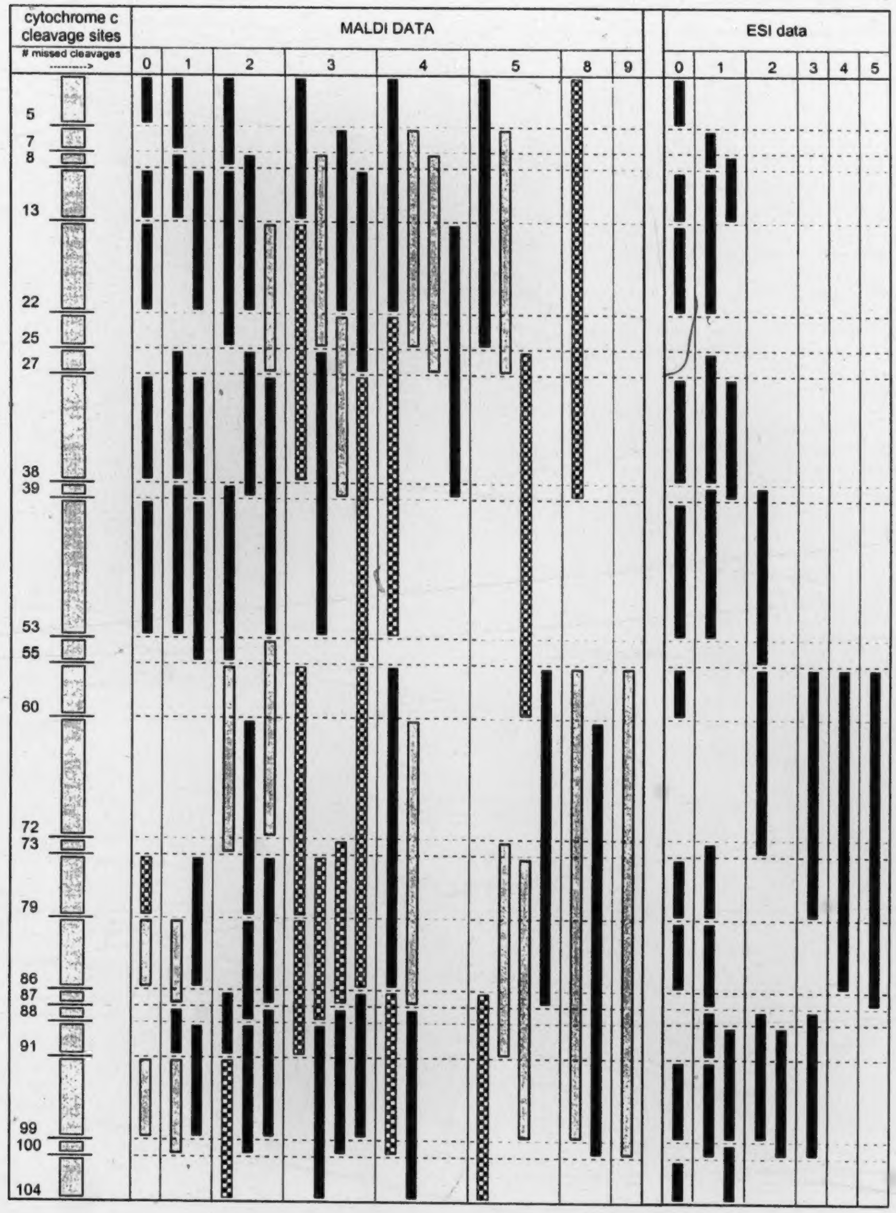


Figure 3.3 Tryptic peptides detected from the digest of cytochrome c bound to the C_{18} surface by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI are indicated with a checkered bar. Fragments commonly observed in each MALDI analysis method is indicated with a black bar.

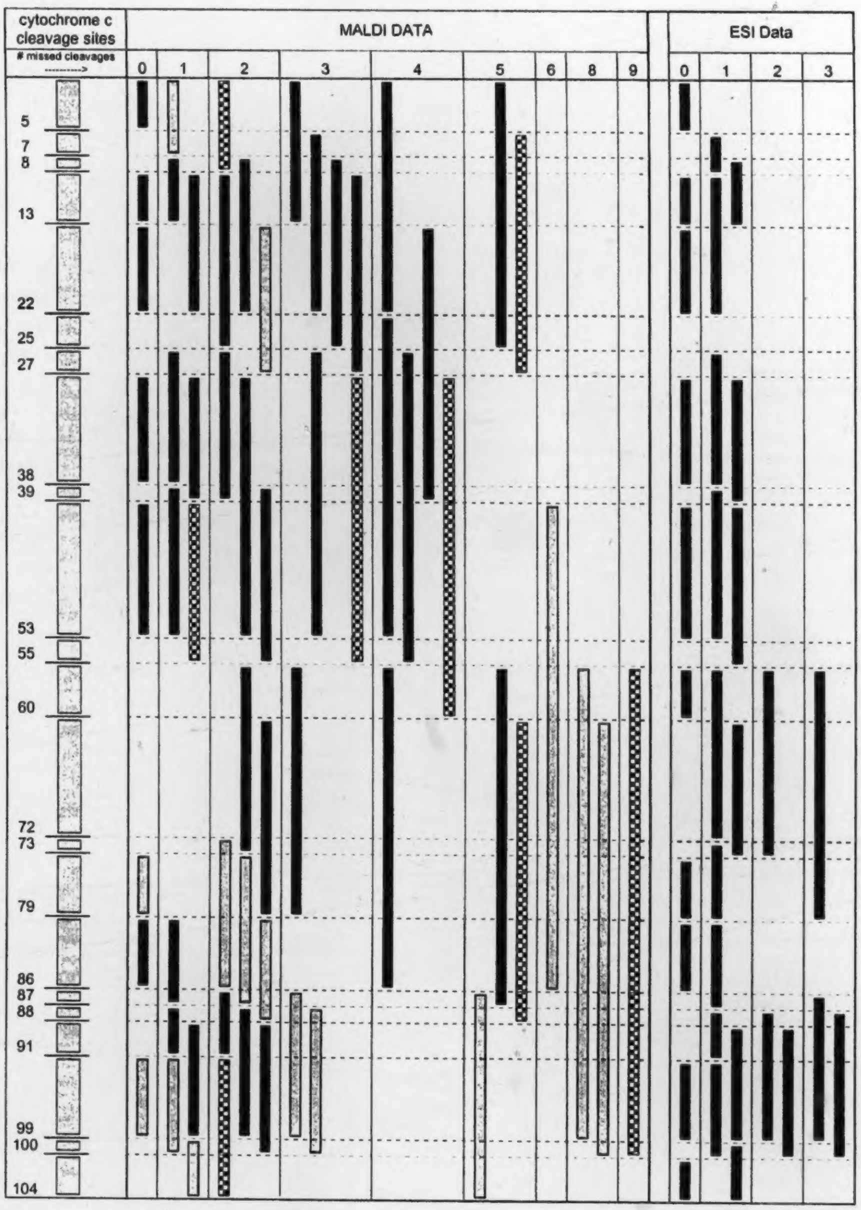


Figure 3.4 Tryptic peptides detected from the digest of cytochrome c bound to the R2 surface by using MALDI and ESI-MS. The fragments observed exclusively in the direct MALDI analysis are indicated with a light gray bar, while those exclusive to LC/offline MALDI are indicated with a checkered. Fragments commonly observed in each MALDI analysis method are indicated with a black bar.

The peptides detected by either MALDI or ESI MS are organized according to the number of missed cleavages they contain.

As Figure 3.2 shows, under the conditions used, a solution digestion of cytochrome c results in a partial digest of the protein, since many incompletely digested peptide fragments were detected. The partial digestion of cytochrome c results in a total of 41 unique peptides observed by MALDI and ESI MS analysis. It is noted that a more complete analysis of the generated peptide fragments is obtained by incorporating all three MS detection schemes since some peptides were exclusively observed in each of the three detection methods. It is likely that additional peptide fragments remained undetected. In particular, short peptide fragments (spanning from 1 to 4 amino acid residues), are difficult to detect by MALDI MS or ESI MS owing to their low molecular weight, or inability to retain on a reversed-phase separation column. Nonetheless, the combination of the three MS detection schemes resulted in 100% sequence coverage for cytochrome c. Also, from the peptide fragments detected as shown in Figure 3.2, all possible tryptic digestion sites of cytochrome c were cleaved and some fragments containing missed cleavage were detected.

Comparing the solution digest (Figure 3.2) to the digests of cytochrome c bound to C_{18} and R2 beads (Figures 3.3 and 3.4), it is obvious that bead digestion generates a larger number of detectable peptide fragments. A total of 66 unique peptide fragments were detected from the R2 bead digest, while 77 unique peptide fragments were observed from the C_{18} bead digest. This is attributed to the fact that the degree of cleavage for each site in the bead digestion is lower than in the solution digestion. Large peptide fragments containing as many as 9 missed cleavages were observed from the bead digests. Most of the larger fragments from the bead digests (having three or more missed cleavages) were not detected from the solution digest. On the other hand, some of the peptides detected in the solution digest having no, or only one missed cleavage, were not detected in the bead digests. In particular, the results of Figures 3.3 and 3.4 reveal a gap in the amino acid sequence between residues 56 to 73 in the MALDI detected peptides with zero or one missed cleavage. Attributed to increased detection sensitivity of lower molecular weight peptides, the ESI results reveal a portion of this sequence (*i.e.* residues from 55 to 66) in the C_{18} digest, but again did not detect the completely digested fragment that spans residues 61 to 72 on either surface. Also, a greater number of peptides having an increased number of missed cleavages were observed by ESI, as well as by MALDI, covering this region of the protein. This therefore indicates that the degree of digestion at this portion of the molecule is significantly reduced by adsorption to the bead surface. Based on the significant number of peptides detected having a greater proportion of missed cleavage sites (>4 per fragment), it can be concluded that the adsorption of cytochrome c to R2 or C_{18} beads results in an overall reduction in digestion efficiency, which is more pronounced at certain regions of cytochrome c.

The combination of the three detection schemes used to analyze the digestion products provides a more detailed investigation on the subtle differences between surface- and solution-phased digests. In particular, analysis by LC/ESI alone would not have revealed most of the larger peptides that were detected by MALDI and LC/MALDI, which constitute the greatest observed differences between the surface and solution digests.

Although less complete digestion is observed as cytochrome c is adsorbed to a solid support, all cleavage sites of cytochrome c were still digested to some degree. Therefore, it can be concluded that protein adsorption does not completely block the digestion of any individual digestion site of this protein. The data shown in Figures 3.3-3.5 also suggest that the effect of support type on the extent of digestion is relatively small compared to the differences resulting between a solution digestion and a digest of a surface-bound protein, which is in agreement with the data from the UV chromatograms.

It should be noted that the observation of bead digestion producing a peptide map containing a much wider range of peptide fragments as compared to solution digestion is significant for protein identification and characterization. Just as various enzymes can be used to generate unique peptide fragments from a protein, this work shows that the adsorption of a protein to a surface can generate unique peptides from a digestion with the same enzyme. Peptide fragments containing a higher degree of missed cleavage sites are useful to confirm the sequence of a segment of protein observed from smaller peptide fragments. In other cases, where cleavage sites are closely spaced within a given protein, segments of a given protein are sometimes only detected with MALDI or ESI as portions of larger fragments with increased missed cleavages. This is particularly useful where the detection of all residues is required, as in the case of amino acid sequencing or the localization of post-translational modifications.

3.3.3 In-Column Digestion of Proteins on Various Supports

As it was demonstrated recently [3], microbeads packed in a microcolumn can be used for protein capture and digestion. Protein capture requires a strong binding of protein to the bead surface, hence favors the use of more hydrophobic surfaces such as C_{18}. However, strong binding may adversely affect the digestion process. The effects of various supports on digestion of proteins captured from a dilute solution are therefore investigated. Figure 3.5 shows the MALDI spectra obtained from the digests of cytochrome c adsorbed to four commonly used chromatographic beads. Solutions consisting of 100 μL of 100-nM protein were used to load the microcolumns. The microcolumns were packed with Poros R2 beads, C_{18} bonded-phase silica beads, C_8 beads, or C_4 beads. The results shown in Figure 3.5 represent the digestion products that were obtained under optimized in-column digestion conditions for the dilute protein solutions used. As seen in Figure 3.5, many peaks corresponding to peptide fragments of cytochrome c are visible in the MALDI spectra (these peaks are labeled "C"). Also, other peaks were detected, including trypsin autolysis fragments (labeled "T" in the figure) as well as a few commonly observed matrix clusters of HCCA. In addition, several unidentified peaks (i.e., most unlabeled peaks in the figure) are visible in the spectra displayed in Figure 3.5. In particular, an increased number of unknown peaks are observed in the MALDI spectrum as the support is varied from R2 to C_{18} to C_8 to C_4. It is suspected that these peaks may also be the fragments of cytochrome c, arising from non-specific tryptic cleavage of the protein while absorbed to the various supports. These unidentified peaks, along with other differences, such as signal intensities, contribute to the observed differences in the MALDI spectra for digests on the various bead supports.

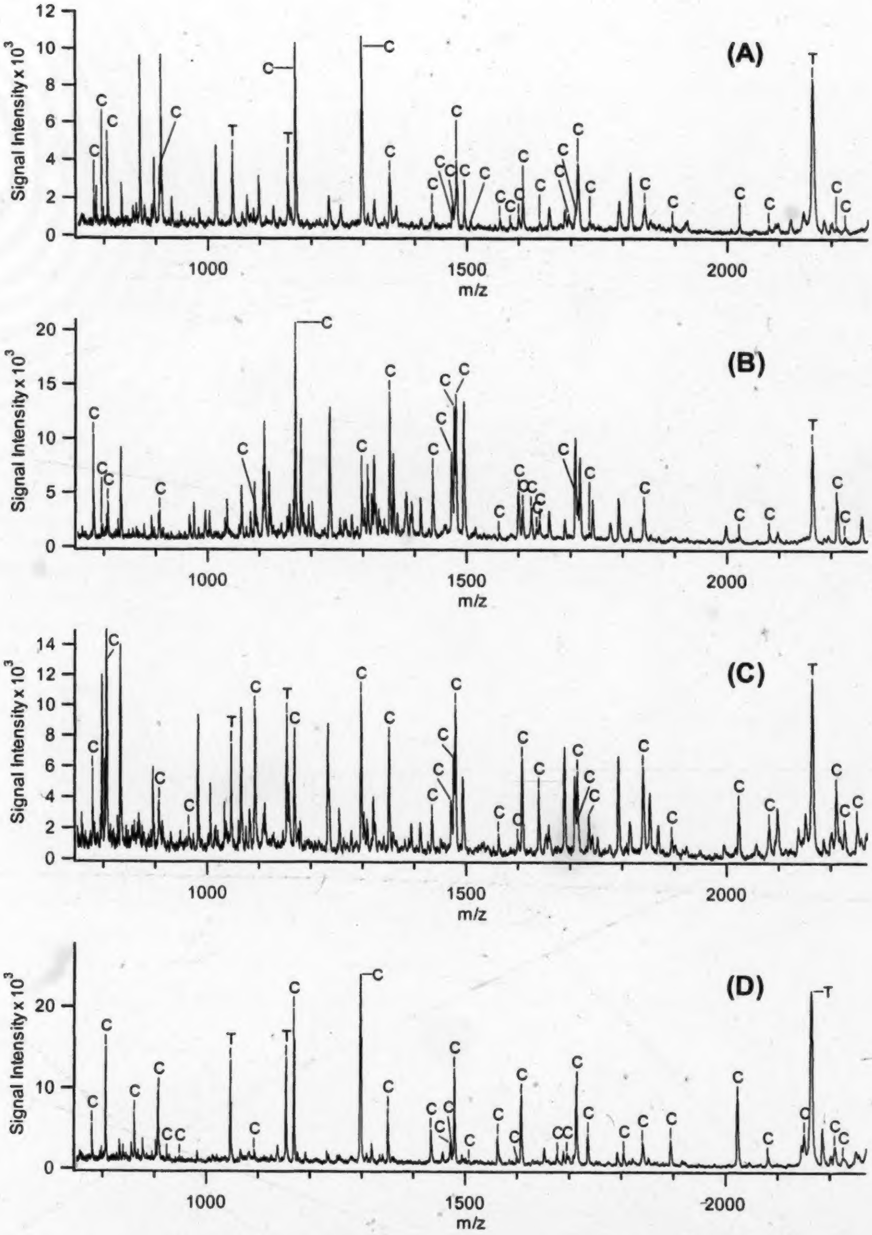


Figure 3.5 MALDI spectra showing the in-column digestion of 100-nM cytochrome c, where the column was packed with (A) Vydac C_{18}, (B) C_8. (C) C_4, and (D) Poros R2 beads. In the spectra, the label "C" refers to peptide fragments of cytochrome c, and "T" refers to peptide fragments resulting from trypsin autolysis.

Although the spectra shown in Figure 3.5 exhibit several differences, the information conveyed by the MALDI spectra for each surface digest is very similar. This is more clearly shown by the data presented in Table 3.1, which lists the peptide fragments detected from each of the four column digestions. From this table, we see that the obtained peptide mass maps from the cytochrome c digests on each type of support are very similar. A similar number of peptide fragments were detected from the digests on each surface type, several of which being common to all four surfaces. Of the peptide fragments that were not commonly observed in each digest, most were of low signal intensity. These results indicate that the choice of surface employed (*i.e.* R2, C_{18}, C_8, or C_4) has little effect on the outcome of the obtained peptide fragments from MS detection.

These results were further confirmed by digestion of various proteins on each of the four support types. A microcolumn digestion of BSA (using 100 μL of 100 nM protein to load the column) also revealed similar results with each of the four surfaces. These results are summarized in Table 3.2. In this case, between 32 and 36 peptides were detected from each individual digest, with 26 peptide fragments being common to each of the four surfaces. Again, the peaks detected that were unique to a single support type were only weakly observed in the MALDI spectra. Similarly, microcolumn digests of 100-nM lysozyme on each of these four surfaces were essentially identical in terms of the number of peaks, as seen in Table 3.2. In this case, between 14 and 17 peaks were detected from each individual digest, and of these 12 peptides were commonly observed in each of the four digests. The 12 commonly observed peptides represent 73% sequence coverage. As high as 100% sequence coverage could be obtained from an individual

Table 3.1 Tryptic peptides detected from the digests of cytochrome c bound to various surfaces.

peak #	amino acid residue	peak mass	Type of hydrophobic surface			
			C_{18}	C_8	C_4	R2
1	1-8	860.5				X
2	1-13	1475.9	X	X	X	X
3	6-13	947.6				X
4	9-22	2250.8			X	
5	14-22	1635.4		X		
6	23-38	1675.9				X
7	23-39	1804				X
8	26-39	1561.9	X	X	X	X
9	26-38	1433.8	X	X	X	X
10	28-38	1168.6	X	X	X	X
11	28-39	1296.7	X	X	X	X
12	39-53	1598.8	X	X	X	X
13	39-55	1840.9	X	X	X	X
14	40-53	1470.7	X	X	X	X
15	40-55	1712.8	X	X	X	X
16	56-72	2081	X	X	X	X
17	56-73	2209.3	X	X	X	X
18	61-72	1495.7	X			
19	61-73	1623.8		X		
20	73-79	806.5	X	X	X	X
21	74-88	1695	X			X
22	80-86	779.4	X	X	X	X
23	80-87	907.5	X	X	X	X
24	87-100	1735	X	X	X	X
25	87-104	2150.2				X
26	88-99/ 89-100	1478.8	X	X	X	X
27	87-99/ 88-100	1606.9	X	X	X	X
28	88-104	2022.1	X	X	X	X
29	89-99	1350.7	X	X	X	X
30	89-104	1894	X		X	X
31	92-99	964.5			X	
32	92-100	1092.6		X	X	X
33	92-104	1507.8	X			X
total number of peaks			23 peaks	22 peaks	23 peaks	28 peaks
sequence coverage			89%	97%	97%	91%

Table 3.2 Tryptic peptides detected from the digests of BSA bound to various surfaces.

peak #	amino acid residue	peak mass	Type of hydrophobic surface			
			C-18	C-8	C-4	R2
1	25-34	1193.6	x	x	x	x
2	29-34	712.4	x	x	x	
3	35-44	1249.6	x		x	x
4	45-88	4910.4			x	
5	66-75	1163.6	x	x	x	x
6	101-105	545.3		x	x	
7	161-167	927.5	x	x	x	x
8	161-168	1083.6	x	x	x	x
9	168-183	2045.0	x	x	x	x
10	198-218	2289.2				x
11	205-211	906.5	x	x	x	x
12	205-222/ 210-228	2144.2		x	x	x
13	219-222	572.4	x	x	x	
14	229-232	508.3	x	x	x	x
15	219-235	2004.1	x	x	x	x
16	223-232	1138.6	x	x	x	x
17	233-241	1001.6	x	x	x	x
18	236-241	689.4	x	x	x	x
19	242-248	847.5	x	x	x	x
20	249-266	2057.2	x			
21	257-266	1153.7	x	x	x	x
22	257-285	3211.6	x		x	x
23	300-309	1177.6		x		
24	347-359	1567.7	x	x	x	x
25	347-360	1723.8	x	x	x	x
26	347-374	3300.8	x	x	x	x
27	360-371	1438.8	x	x	x	x
28	361-371	1283.7	x		x	x
29	400-420	2539.3	x	x	x	x
30	400-433	4000.1	x	x	x	x
31	402-412	1305.7	x	x	x	x
32	421-433	1479.8	x	x	x	x
33	421-436	1900.0	x	x	x	x
34	437-451	1639.9	x	x	x	x
35	437-455	2025.2		x		
36	437-459	2438.4		x		x
37	438-451	1511.8		x		x
38	452-459	817.5		x		
39	452-482	3515.7				x
40	548-557	1142.7	x	x	x	x
41	549-557	1014.6	x	x	x	x
42	549-561	1504.9		x		
43	549-587	4407.2			x	x
44	558-568	1308.7	x	x	x	x
45	569-580	1399.7				x
46	569-597	3153.4			x	
total number of peaks			32 peaks	36 peaks	36 peaks	36 peaks
sequence coverage			38%	36%	47%	48%

digest, as some large individual fragments can contribute to a somewhat greater difference in the sequence coverage.

I note that I have not investigated the effect of surface chemistry on digestion of surface-bound very hydrophobic proteins. It was shown earlier [6] that this type of protein requires chemical cleavage prior to trypsin digestion.

3.3.4 Effects of Bead Construction and Pore Size on Digestion

I have explored the possible effects of pore size of the beads on protein digestion. Proteins may enter pores of the beads, making them less accessible to trypsin. To address this, I compared the peptide maps obtained from the digestion of cytochrome c on various polystyrene divinylbenzene surfaces, each having different pore sizes. The four different beads used in this study included PRP-3 beads (mean pore size 300 Å, 12–20 μm diameter), PRP-1 beads (mean pore size 100 Å, 12 – 20 μm diameter), and PRP-Infinity beads (non-porous, 4 μm diameter), as well as R2 beads (having through-pores of 6000-8000 Å, and diffusive pores of 800-1500 Å, bead diameter 20 μm). Both the through pores and diffusive pores are significantly larger than typical proteins. The obtained MALDI spectra from digests of cytochrome c, lysozyme, and BSA adsorbed on these beads were all similar in terms of the number and type of peptides detected (data not shown). In particular, samples digested on the PRP-Infinity beads, which do not have any pores, did not improve the quality of the MALDI spectrum, compared to samples digested on beads of varying pore sizes. Therefore, it is concluded that pore size does not have a significant effect on the digestion efficiency of an adsorbed protein.

3.3.5 Enzymatic and Chemical Cleavage of Proteins Adsorbed on Microbeads

Trypsin digestion on different surfaces has been thoroughly studied. However, other enzymes and reagents are also commonly used for protein degradation. Thus, the enzymes chymotrypsin and pepsin, along with CNBr, were used to fragments proteins. Chymotrypsin and pepsin were used to digest BSA and cytochrome c respectively. CNBr was used to chemically cleave β-Casein. Digestion was performed on beads and compared to digests performed in solution. Chymotrypsin was first used because it is similar to trypsin. The digestion conditions used were identical to typsin however, chymotrypsin cleaves at the C-terminal side of phenylalanine, tyrosine, tryptophan and leucine. A 100 μL solution of 500 nM BSA was adsorbed to a 30 μg Poros R2 mico column. The protein was digested with chymotrypsin at pH 8 for 2 hours at 37 °C. Following digestion the beads were washed with water and placed directly onto a MALDI target for analysis. The digestion of BSA on microbeads with chymotrypsin was successful. Figure 3.6 displays the MALDI spectrum from the peptide fragments and Table 3.3 summarizes the data. The digests in solution and on beads have similar MALDI spectra. There were 18 peaks, which were common between both spectra.

The second enzyme used was pepsin. Pepsin was chosen because it preferentially cleaves at hydrophobic amino acids; phenylalanine, leucine, and glutamic acid. It does not cleave at valine, alanine or glycine. Other residues may be cleaved, with very variable rates. Since these amino acids are hydrophobic, they should be closer to the bead and less accessible by enzymes. It has been reported that pepsin will not digest proteins adsorbed

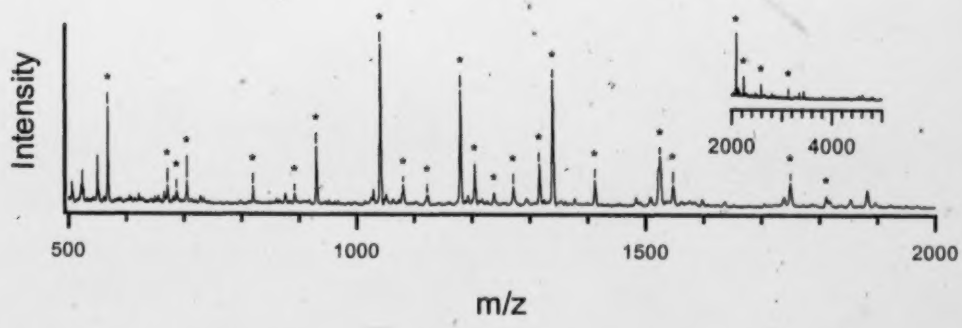


Figure 3.6 MALDI spectra showing the chymotrypsin digestion of 500-nM BSA on Poros R2 beads packed in a column. In the spectra, the label "*" refers to peptide fragments of BSA. Insert displays MALDI spectra from 2000 Da to 5000 Da.

Table 3.3 Summary of different enzymes and chemicals used for degradation of proteins to peptides.

Protein	Enzyme or Chemical Cleaving Agent	Digestion	# matching peaks (*)	Amino Acid % Coverage
BSA	Chymotrypsin	Solution	29	51
		Beads	25	40
Cytochrome c	Pepsin	Solution	14	98
		Beads	10	72
β-Casein	CNBr	Solution	7	100
		Beads	7	100

on the beads because these cleavage sites are inaccessible [18]. Pepsin digestion conditions require a pH of 2 and digestion time of 1 hour at room temperature. Aliquots of 100 μL containing 500 nM cytochrome c were digested in solution and on beads. Figure 3.7 shows the MALDI spectrum of the peptide fragments generated from digesting cytochrome c with pepsin on beads and Table 3.3 summarizes the comparative results. The MALDI spectra revealed similar peptide mass maps. There were 4 peaks that were common between both spectra.

The third experiment used a chemical to degrade a protein. Cyanogen bromide (CNBr) was chosen since it is commonly used to fragment hydrophobic proteins. CNBr cleaves at the methionine amino acid and is very specific. The digestion of β-Casin on microbeads with CNBr was successful. Figure 3.8 displays the MALDI spectrum of the digest on beads. The digests in solution and on beads have similar MALDI spectra. There were 14 peaks detected in both the solution and bead digests. All of the peaks were common in both spectra.

Other enzymes were used for protein degradation: Elastase, Arg-c, Lys-c, and Glu-c (data not shown). All enzymes successfully digested proteins on beads. Therefore this method is not limited to the type of enzyme need to digest a particular protein.

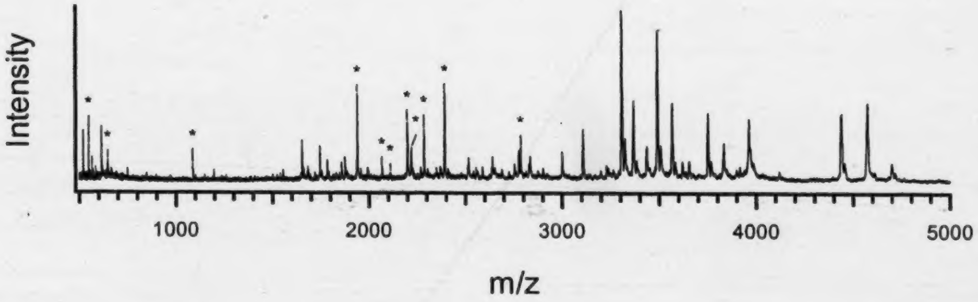


Figure 3.7 MALDI spectra showing the pepsin digestion of 500-nM cytochrome c on Poros R2 beads packed in a column. In the spectra, the label "*" refers to peptide fragments of cytochrome c.

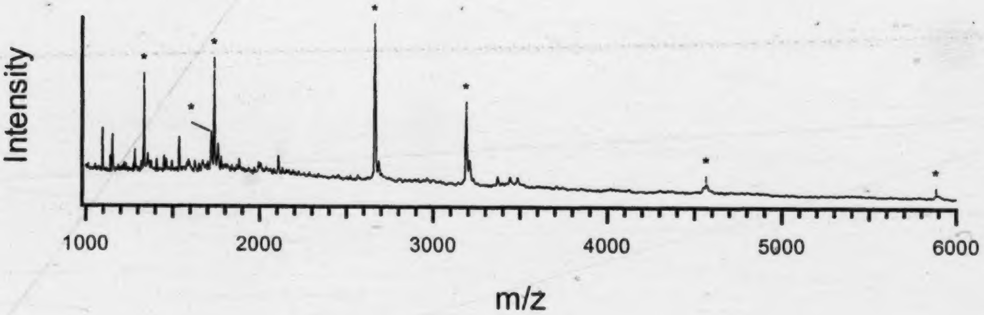


Figure 3.8 MALDI spectra showing the CNBr cleavage of 500-nM β-Casein on Poros R2 beads packed in a column. In the spectra, the label "*" refers to peptide fragments of β-Casein.

3.3.6 Digestion of Hydrophobic Protein

The proteome of a cell contains proteins of varying hydrophobicity. Many hydrophobic proteins are of particular importance since they constitute membrane proteins that are involved in transporting and signaling pathways. Here, I examine the performance of the in-column digestion procedure as applied to hydrophobic proteins. Bacteriorhodopsin was chosen as a model hydrophobic protein to demonstrate the performance of the method. A 100-μL aliquot of 500-nM bacteriorhodopsin, prepared by dilution of the solid protein in water, was injected into a column packed with 30 μg of R2 beads. During sample loading, the protein was seen to collect as a purple precipitate at the head of the microcolumn. It then became clear that even at this low concentration, the highly hydrophobic bacteriorhodopsin did not solubilize in aqueous solution. An in-column tryptic digestion of the precipitate yielded no significant peptide fragments in the MALDI spectrum.

The incorporation of a cyanogen bromide (CNBr) digestion serves to solubilize the protein, or fragments of the protein, and therefore makes subsequent enzyme digestion with trypsin possible. Therefore a procedure was developed for in-column digestion of bacteriorhodopsin with CNBr. Following sample loading of 500-nM bacteriorhodopsin on the column, a 40 μL solution of 50 mM CNBr, prepared in 12.5% aqueous TFA, was passed through the column. The microcolumn was left overnight to allow for digestion. Following the CNBr digest, the column was washed with 50 mM NH_4HCO_3 until a pH of 8 was established in the column, then the buffered trypsin solution was passed through the column and incubated for 30 min. The peptides were then eluted and analyzed by MALDI. The MALDI spectrum from the analysis of the peptide fragments is displayed in Figure 3.9A. A total of 9 peaks were detected corresponding to bacteriorhodopsin peptide fragments, giving 19% sequence coverage. A comparative solution-phase digestion of 500-nM bacteriorhodopsin with CNBr followed by trypsin revealed no detectable peptide fragments for the hydrophobic protein. At a concentration of 2 μg/μL (74 μM) a MALDI spectrum comparable to the in-column digest at 500 nM was achieved. The MALDI spectrum from the solution-phase digestion of this sample is shown in Figure 3.9B. Therefore it is evident that the combined CNBr/trypsin digestion in a column is an effective method for generating peptides from dilute hydrophobic protein samples.

3.4 Conclusions

Protein samples that are soluble in aqueous solution can be digested with trypsin while adsorbed on hydrophobic surfaces, namely reversed-phase chromatography beads. UV chromatographic analysis of the protein digests clearly indicates that the overall digestion process is altered in going from a solution digest to a surface bound digestion. However, only minor differences are observed upon varying the surface from R2 to C_{18} beads. Mass spectrometric analysis specifically shows the difference in the type of peptides generated from these two methods (solution *vs.* surface-bound digests). I demonstrate that using a combination of direct MALDI, LC/offline MALDI, and LC/ESI MS analysis, instead of using a single MS detection technique alone, subtle changes in protein digestion characteristics can be revealed. It is shown that a greater number of peptides containing missed cleavage sites are detected in the surface- digestion.

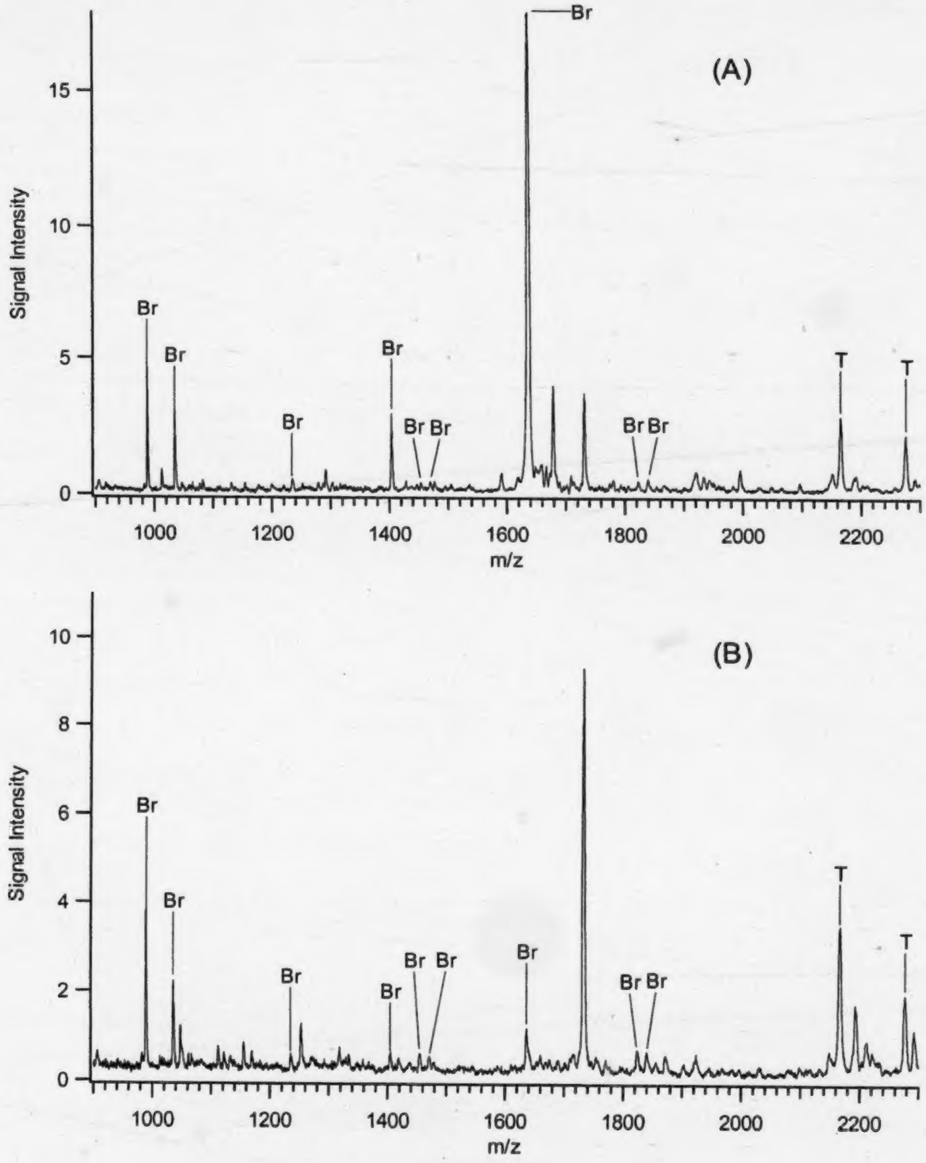


Figure 3.9 MALDI spectra showing the products of (A) in-column digestion of 500 nM bacteriorhodopsin by a combined CNBr/trypsin and (B) in-solution digestion of 2 μg/μL bacteriorhodopsin with CNBr/trypsin. Br = peptide fragments of bacteriorhodopsin, T = peptide fragments from trypsin autolysis.

I have also demonstrated that peptide mass maps obtained by direct MALDI analysis show similar levels of protein sequence coverage for the digests of proteins bound to C_{18}, C_8, C_4, and R2 beads. In addition, the pore size of the beads does not have a significant effect on the digestion characteristics of an adsorbed protein. Thus, for protein identification by peptide mass mapping using direct MALDI analysis of the surface-bound protein digest, any one of the hydrophobic surfaces examined herein can be used. The choice of the micro-bead surface is therefore mainly dependent on the efficiency of protein retention, which can be optimized for different applications. For example, C_{18} beads should provide a better retention for many proteins and thus they are the preferred media for routine micro-column digestion experiments.

Digestion on beads was achieved with chemicals and an array of enzymes. This is important since not all proteins are easily cleaved with trypsin. Some proteins require other enzymes in order to create a thorough peptide mass map. Hydrophobic proteins can also be digested while adsorbed in a microcolumn. In this case, in-column digestion of the protein with CNBr prior to tryptic digestion aids in protein solubilization, and therefore increases the number of detectable peptide fragments.

3.5 Literature Cited

1. Aebersold, R. H.; Leavitt, J.; Saavedra, R. A.; Hood, L. E.; Kent, S. B. *Proc. Natl. Acad. Sci. USA* **1987**, *84*, 6970-6974.
2. Reid, G.; Gan, B. S.; She, Y. M.; Ens, W.; Weinberger, S.; Howard, J. C. *Appl. Environ. Microbiology* **2002**, *68*, 977-980.
3. Issaq, H. J.; Veenstra, T. D.; Conrads, T. P.; Felschow, D. *Biochem. Biophys. Res. Commun.* **2002**, *292*, 587-592.
4. Merchant, M.; Weinberger. S. R. *Electrophoresis* **2000**, *21*, 1164-1171.
5. Doucette, A.; Craft, D.; Li, L. *Anal. Chem.* **2000**, *72*, 3355-3362.
6. Craft, D.; Doucette, A.; Li, L. *J. Proteome Research*, **2002**, *1*, 537-547.
7. Craft, D.; Doucette, A.; Li, L. In *Proceedings of the 48th ASMS Conference on Mass Spectrometry and Allied Topics*, Long Beach CA, 2000, p. 1429.
8. Vestling, M. M.; Fenselau, C. *Anal. Chem.* **1994**, *66*, 471-477.
9. Bienvenut, W. V.; Sanchez, J. C.; Karmime, A.; Rouge, V.; Rose, K.; Binz, P.-A.; Hochstrasser, D. F. *Anal. Chem.* **1999**, *71*, 4800-4807.
10. Courchesne, P. L.; Luethy, R.; Patterson, S. D. *Electrophoresis* **1997**, *18*, 369-381.
11. Lui, M.; Tempst, P.; Erdjument-Bromage, H. *Anal. Biochem.* **1996**, *241*, 156-166.
12. Fabris, D.; Vestling, M. M.; Cordero, M. M. Doroshenko, V. M.; Cotter, R. J.; Fenselau, C. *Rapid Comm. Mass Spectrom.* **1995**, *9*, 1051-1055.
13. Vestling, M. M.; Fenselau, C. *Mass Spectrom. Rev.* **1995**, *14*, 169-178.
14. Schleuder, D.; Hillenkamp, F.; Strupat, K. *Anal. Chem.* **1999**, *71*, 3238-3247.

15. McComb, M. E.; Oleschuk, R. D.; Manley, D. M.; Donald, L.; Chow, A.; O'Neil, D. J.; Ens, W.; Standing, K. G.; Perreault, H. *Rapid Comm. Mass Spectrom.* **1997**, *11*, 1716-1722.

16. McComb, M.; Oleschuk, R. D.; Chow, A.; Ens, W.; Standing, K. G.; Perrault, H.; Smith, M. *Anal. Chem.* **1998**, *70*, 5142-5149.

17. Lee, C. H.; McComb, M. E.; Bromirski, M.; Jilkine, A.; Ens, W.; Standing, K. G.; Perrault, H. *Rapid Comm. Mass. Spectrom.* **2001**, *15*, 191-202.

18. Aguilar, M.; Clayton, D. J.; Holt, P.; Kronina, V.; Boysen, R. I.; Purcell, A. W.; Hearn, M. T. W. *Anal. Chem.* **1998**, *70*, 5010-5018.

19. Jemmerson, R.; Paterson, Y. *Science* **1986**, *232*, 1001-1004.

20. Suckau, D.; Kohl, J.; Karwath, G.; Schneider, K.; Casaretto, M.; Bitter-Suermann, D.; Przybylski, M. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 9848-9852.

21. Macht, M.; Fiedler, W.; Kurzinger, K.; Przybylski, M. *Biochemistry* **1996**, *35*, 15633-15639.

22. Burnens, A.; Demotz, S.; Corradin, G.; Binz, H.; Bosshard, H. *Science* **1987**, *235*, 780-783.

23. Schreimer, D.C.; Li, L. *Anal. Chem.* **1996**, *68*, 3382-3378.

24. Schreimer, D.C.; Yalcin, T.; Li, L. *Anal. Chem.* **1998**, *70*, 1569.

25. Litborn, E.; Emmer, A.; Roeraade, J. *Anal. Chim. Acta* **1999**, 401, 11-19.

26. Kosciarz E.; Vercaigne-Marko D.; Scellier N.; Nedjar-Arroume N.; Guillochon D. *Biotechnol. Appl. Biochem.* **1999**, *30*, 147-153.

27. Vercaigne-Marko D.; Kosciarz E.; Nedjar-Arroume N.; Guillochon D. *Biotechnol. Appl. Biochem.* **2000**, *31*, 127-134.

28. Gobom, J.; Nordhoff, E.; Ekman, R.; Roepstorff, P. *Int. J. Mass Spectrom. Ion. Processes* **1997**, *169/170*, 153-163.

29. Annan, R. S.; McNulty, D. E.; Carr, S. A. In *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics*, Portland OR, 1996; p. 702.

30. Dai, Y.; Whittal, R. M.; Li, L. *Anal. Chem.* **1999**, *71*, 1087-1091.

31. Whittal, R. M.; Li, L. *Anal. Chem.* **1995**, *67*, 1950-1954.

Chapter 4

An Automated Approach to Surface Digestion

4.1.1 Introduction

The current method of choice for proteome analysis is the combination of two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) for protein separation with mass spectrometry for protein identification [1-6]. 2D-PAGE possesses several desirable properties. The gels deliver accurate and detailed information on a particular protein including its relative abundance, potential post-translational modifications, isoelectric point (pI), and molecular weight [7-10]. The true power of this technique is its ability to separate thousands of protein in a single gel [11]. However, the dynamic range of protein visualization, and the requirement that the protein be soluble in aqueous solution without detergent in the first dimension of separation complicates the detection and separation of low-abundance and hydrophobic proteins by 2D-PAGE [12]. Also, the efficiency of in-gel protein digestion, as well as extraction of the resulting peptides can further reduce the sensitivity of protein identification. In addition, this method proves to be very labour intensive. The strategy of (1) separating proteins by 2D-PAGE, (2) locating protein spots in the gel by staining, (3) cutting out spots of interest, (4) digesting the excised protein spots, (5) extracting the peptide fragments from the gel, and (6) analyzing the tryptic digests by mass spectrometry has been automated through a series of robotic operations [13-18]. However, this approach is still very complex and does not interface easily with mass spectrometry. These drawbacks have stimulated the development of alternative technologies to the traditional approach for protein identification [19].

Various techniques have been used to manipulate protein samples prior to their introduction into the mass analyzer. For example, the isolation and/or separation of proteins in a sample could be achieved by solvent extraction, chromatography such as reversed-phase liquid chromatography, or by gel electrophoresis. The pre-concentration of proteins in a sample is often achieved through solvent evaporation, precipitation, affinity capture, or adsorption of the protein to membranes or small particles [20-22]. Sample cleanup can be achieved through extraction, membrane filters, molecular weight cut-off filters, or chromatography. In addition, several procedures have been developed which focus on improving digestion efficiency. Procedures have been developed involving the use of immobilized enzymes, such as trypsin, to digest proteins in a flow-through fashion [23], or to digest very small (sub-microliter) sample volumes by digestion in nanovials [24], capillaries [25], or in etched channels on chips [26]. In cases where more than one manipulation step is required, such as sample cleanup, pre-concentration, as well as digestion, a combination of several of the aforementioned techniques can be employed. However, the use of multiple steps will increase the difficulty of sample preparation, requiring more time and effort. Also, with each additional step comes the possibility of sample loss or contamination of sample, which further complicates the sample manipulation step. Most existing automated technologies for performing sample manipulations and modifications focus on perhaps one or two sample manipulation/modification steps at a time. In order to perform all required modifications, these conventional methods are performed in succession. Several devices have been reported for automation of the conventional methods; however, few methodologies are available which truly combine all manipulations in a single device [27,28].

In Chapter 2, a technique was described which allowed for the capture of proteins in a dilute solution onto a reversed-phase column. The bound proteins were then subjected to trypsin proteolysis and the peptides were eluted from the column for identification by mass spectrometry. This approach was found to be particularly useful in handling dilute solutions of protein present in a sample. This technique clearly demonstrated the advantages of digesting on a column. However, for many reasons, it would not be practical to perform this procedure on a complex protein mixture, as required in proteome analysis. The columns were not built for multiple uses. Therefore each sample required a new column. The samples were loaded, digested, and eluted manually with a syringe pump. The columns were packed in a gel loader tip, which could not be directly coupled to a mass spectrometer. The demands of proteomics require a system capable of complete sample manipulation. In order to handle the growing number of samples needing analysis, this system must be automated and directly coupled to a mass spectrometer.

In the work presented herein, a microcolumn is designed to meet the demands of high throughput proteome analysis. The method of protein digestion with this column makes use of an automated HPLC system to manipulate samples, and is directly coupled with a mass spectrometer. The system was optimized with the analysis of several standard proteins. Using the device, the labour-intensive nature of high throughput applications is significantly reduced, allowing for rapid analysis of a multitude of samples with very little effort.

4.2 Experimental

4.2.1 Materials and Reagents

Equine cytochrome *c*, bovine serum albumin (BSA), and trypsin (from bovine pancreas, TPCK treated to reduce chymotrypsin activity; dialyzed, lyophilized) were from Sigma Aldrich Canada (Oakville, ON). Poros R2 beads (20 μm) were from PerSeptive Biosystems (a gift of Professor F. Cantwell, University of Alberta). All PEEK tubing and unions were purchased from Upchurch (Oak Harbor, WA, USA). A 2-μm stainless steel frit from Upchurch Scientific (catalogue number C-407x) was used in the microcolumns. Analytical grade acetone, methanol, acetonitrile, acetic acid, and trifluoroacetic acid (TFA) were purchased from Caledon Laboratories (Edmonton, AB,). Water used in the experiments was from a NANOpure water system (Barnstead/Thermolyne). All other chemicals were purchased from Sigma Aldrich Canada.

4.2.2 Microcolumn Design

The R2 Poros beads were packed inside 1/16-inch PEEK tubing, having an i.d. of 254 μm. Prior to packing a column, the beads were suspended in methanol at a concentration of 10 mg/mL. A 4 μL aliquot was injected into the column resulting in 40 μg of packing. A stainless steel frit with a pore size of 2 μm was used to contain the beads. The outside of the column was wrapped with heating wire and insulated with a cotton wrap. A temperature sensor was placed between the heating wire and column, and the temperature was set to 37°C with a Variac voltage regulator. A schematic representation of the column is displayed in Figure 4.1.

4.2.3 Sample Manipulation

4.2.3.1 Loading and Washing

The primary step in the automation process involves capturing the proteins on the column packing material. Prior to sample injection, the microcolumn was first equilibrated in pure water at a flow rate of 10 µL/min for 5 min. The protein sample was then injected into the microcolumn using the auto-sampler of the Agilent HPLC system and ChemStation software. The sample was injected at a flow rate of 5 µL/min. Typical sample volume for injection was 100 µL. Therefore, 30 min was allowed for the protein to reach the microcolumn and adsorb. Following sample injection the column was flushed with water for an additional 5 min at 10 µL/min. The entire loading and washing process requires 35 min. Once the protein was bound and free of all contaminates it was ready for digestion.

4.2.3.2 On-line Protein Digestion

A 31 ng/µL typsin solution in water, and a 0.2 M ammonium bicarbonate buffer were placed in separate vials of the autosampler. Mixing of trypsin and bicarbonate buffer was accomplished with the HP1100 autosampler. The auto sampler was also set at 4°C to limit trypsin autolysis in pure water. A 10-µL plug of buffered trypsin was injected at 10 µL/min until the centre of the plug reached the microcolumn. This varied

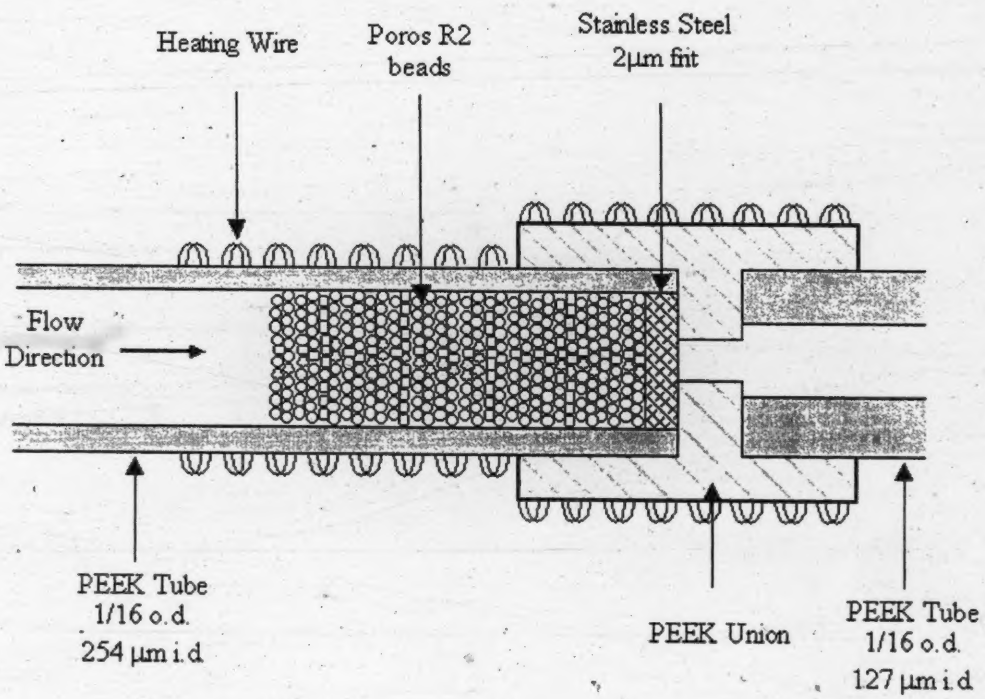

Heating Wire
Poros R2 beads
Stainless Steel 2μm frit
Flow Direction
PEEK Tube 1/16 o.d. 254 μm i.d
PEEK Union
PEEK Tube 1/16 o.d. 127 μm i.d

Figure 4.1 Illustrated representation of micro column used for digestion. The column contains 40 μg of Poros R2 beads. The heating wire was heated to 37°C.

from 2 to 3 min depending on which column was used. The flow was then stopped for 30 min to allow for proteolysis. Following the digestion the column was washed with pure water for an additional 5 min at a flow rate of 10 μL/min. As soon as the excess trypsin and ammonium bicarbonate were washed away, switching valves redirected the flow to the C_8 separation column and mass spectrometer.

4.2.3.3 Elution of Peptide Fragments

With the proteins reduced to peptides and all the excess trypsin and buffers removed, the peptides were ready to be eluted from the microcolumn onto a C_8 reversed-phase column for separation. The HPLC pump used for elution was a Finnigan Surveyor (San Jose, CA, USA). A splitter was introduced prior to the switching valve in order to achieve a flow rate of 1 μL/min to the mass spectrometer. The column was a Vydac reversed-phase (5 μm) n-octylsilica column (Anaheim, CA, USA) with dimensions of 150 μm × 15 cm. A gradient separation was performed using water with 0.5% acetic acid (solvent A) and acetonitrile with 0.5% acetic acid (solvent B). The gradient used was as follows: increase B from 5 to 10% over 2 min, then from 10 to 45% B over 38 min, then increase B to 65% over 10 min. The peptides, once eluted, were detected on a Finnigan LCQ^{Deca} ion trap mass spectrometer (San Jose, CA, USA).

4.2.4 Configuration of On-line Digestion System

Two micro digestion columns, two C_8 columns and three HPLC pumps were coupled to a mass spectrometer. The set-up of the system is schematically depicted in

Figure 4.2. Switching valves were used to divert the flow of each HPLC pump to the appropriate columns. Pump 1 contains an auto-sampler and was exclusively used for loading, digesting, and washing the microcolumns. Switching valve 1 diverted the flow from this pump to one of the digestion columns. Pump 2 has a splitter to achieve a low flow rate of 1 μL/min. This pump was used for eluting peptides from the digestion column into the C_8 column and mass spectrometer. Switching valve 1 diverted the flow from the pump to the other digestion column. The second switching valve was used to select the C_8 column. The third switching valve was used to divert the eluent from the C_8 column into the mass spectrometer. Pump 3 also contains a splitter in order to wash and equilibrate the second C_8 column. Switching valve 2 was used to direct the flow of this pump to the second C_8 column. Figure 4.2 illustrates how pump 1 loads and digests microcolumn 1, while pump 2 elutes and washes microcolumn 2 with the second C_8 column. The time required for loading, digesting, and equilibrating microcolumn 1 was equivalent to the time required to elute peptides and wash the second microcolumn and the second C_8 separation column. The time required for each cycle was 80 min. Once the cycle was completed the switching valves were switched for the next sample. The first switching valve diverted pump 1 to the second microcolumn and pump 2 was diverted to microcolumn 1. The second and third switching valves selected the other C_8 separation column. Table 4.1 summarizes the switching valve positions required for each cycle, while Table 4.2 displays the gradients used for each HPLC pump.

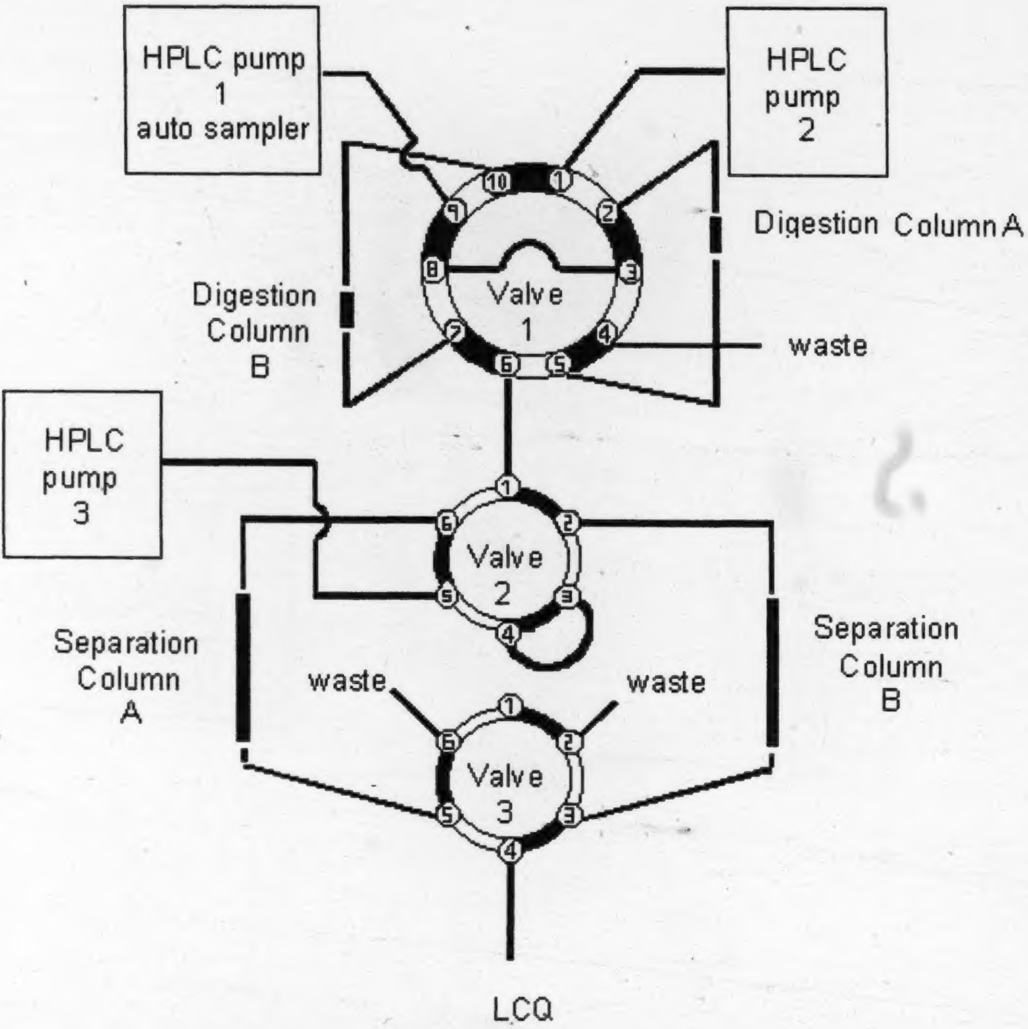

HPLC pump
1
auto sampler
HPLC
pump
2
Digestion Column A
Digestion
Column
B
Valve
1
waste
HPLC
pump
3
Valve
2
Separation
Column
A
Separation
Column
B
waste
waste
Valve
3
LCQ
1
2
3
4
5
6
7
8
9
10

Figure 4.2 Schematic representation of the on-line sample manipulation workstation. If all three switching valves are diverting flow in the black position then HPLC pump 1 is loading and digesting micro column A. HPLC pump 2 is eluting peptides from micro column B into separation column B and into the LCQ^{DECA} for detection. The third HPLC pump is washing separation column A. Once the cycle is completed all three switching valves are changed to the white position.

Table 4.1 Summary of switching valve positions for each cycle of the system.

Cycle	Switch Valve 1	Switch Valve 2	Switch Valve 3	Dig. Col. 1	Dig. Col. 2	Sep. Col. 1	Sep. Col. 2
1	Black	Black	Black	Load + Digest	Elute	Wash	Elute
2	White	White	White	Elute	Load + Digest	Elute	Wash

Table 4.2 Summary of gradients used for each HPLC pump.

Pump 1	Pump 2		Pump3	
Equilibrate 5 min	Gradient Elution		Wash and Equilibration	
Load 30 min	Time	%B	Time	%B
Wash 5 min	0	5	0	5
Load enzyme 5 min	2	10	30	80
Digest 30 min	40	45	40	0
Wash 5 min	50	65	80	0
Total time: 80 min	60	80	Total time: 80 min	
	Total time: 80 min			

4.2.5 Instrumentation

Proteins were loaded onto the micro digestion columns with an Agilent 1100 chromatographic system (Palo Alto, CA, USA) equipped with an autosampler. A second Agilent 1100 pump was used to wash one of the Vydac n-octylsilica (C_8) columns (Anaheim, CA, USA). There are two C_8 columns, each with the dimensions of 150 µm × 15 cm. A third pump, Thermo Finnigan Surveyor (San Jose CA, USA), was used to elute peptides from the micro and C_8 columns into the mass spectrometer. Switching valves used in the automated system were from Rheodyne (Rohnert Park, CA, USA). A Finnigan LCQ^{Deca} ion trap mass spectrometer (San Jose, CA, USA) was used to detect and perform MS/MS on the peptides. Data analysis was performed by SEQUEST.

4.2.6 SEQUEST Analysis

SEQUEST was used to determine the amino acid sequence and thus the protein(s) that correspond to the mass spectrum being analyzed. Each sample was run against the nr.fasta (non-redundant) database from the National Centre for Biotechnology Information (Bethesda, MD, USA). This database was indexed for speed using the parameters; trypsin as the enzyme, and 3 missed cleavages. The Xcorr (cross-correlation) values were used as a primary indicator for identifying proteins. An Xcorr value of 2.0 or greater was used to identify proteins. This is in agreement with Yates *et al.* [29] (see Appendix A for definitions of each SEQUEST parameter).

4.3 Results and Discussion

4.3.1 Optimization of System

In Chapter 3 different reversed-phase material used for capturing and digesting proteins were characterized. It was concluded that only minor differences were apparent when comparing C_4, C_8, C_{18}, and R2 beads. Since there were only minor differences in the digestion efficiency on these materials, R2 beads were selected based on their robustness. It is imperative that the beads are durable in order to achieve automation because each column must be re-used numerous times. The column will also be subjected to both acidic and basic conditions as well as a variety of salts and buffers. Therefore a polymeric surface is preferred over a silica support and thus, R2 beads were used. In Chapter 2, I discussed the importance of column size. If the column is too small, not enough protein is captured and the peptides become difficult to detect. If the column is too large, trypsin will bind to the column and the MS spectrum will be dominated by tryptic peptides. For this present work, 4 μg of Poros R2 beads were used per picomole of protein.

In Chapters 2 and 3, the microcolumns used for digestion consisted of a packed gel loader tip. In order to couple the column with a mass spectrometer the column had to be redesigned. The first column design comprised two glass capillaries, one inside the other. One glass capillary had an o.d. of approximately 5 μm smaller then the i.d. of the larger capillary. A glass fibre membrane was used as a filter. Unfortunately the glue used to hold the two capillaries could not withstand the high pressure of HPLC. Also, the glass fibre membranes were not suited for multiple runs. Therefore, in the subsequent experiments, polyetheretherketone (PEEK) tubing was employed for the column. In addition, the glass fibre membranes were not suitable for multiple runs. Fortunately, stainless steel fibre membranes with 2 μm pores were are available, which were not only durable but also fit inside a PEEK union. Thus, a durable column capable of multiple runs, wide pH range tolerance, and resistant to high pressures associated with HPLC was created.

The entire system relies on the efficiency of digesting proteins in the column. In chapter 2, digestion of proteins on R2 beads in a column by introducing a plug of buffered trypsin was achieved. Since this method proved to be successful, the same procedure was incorporated in the automated system. The introduction of trypsin could be achieved by placing a vial of buffered trypsin in the auto sampler and injecting a 10-μL plug into the microcolumn. However, since the goal of automation would involve setting up a sequence of samples, the trypsin solution would be left in a buffered solution for a significant amount of time. During this waiting period, trypsin autolysis would occur and render the trypsin inactive. For this reason 31 ng/μL typsin and 10 μL of 0.2 M ammonium bicarbonate were place in separate vials. The number of ammonium bicarbonate vials equalled the number of samples being run. Only one vial of trypsin was needed. The injector on the Agilent 1100 system was programmed to withdraw 10 μL from the trypsin vial and dispense it into one of the ammonium bicarbonate vials. To better mix the solution, the injector was programmed to withdraw and eject 10 μL of the solution in the vial. Finally 10 μL was withdrawn from the vial and injected into the system. The autosampler was set at 4°C to limit trypsin autolysis in pure water. As a result active trypsin was introduced into the microcolumns for each sample and efficient digests ensued.

In general, salts and/or buffers are used to isolate proteins and are therefore present in the sample. As mentioned in previous chapters, buffers and salts are unwanted during the digestion process. Therefore, a wash step was incorporated following the loading of a protein sample. Following an on-column digest, unwanted excess trypsin and ammonium bicarbonate are present in the column. Therefore it was necessary to wash the microcolumns after the digestion stage as well. The optimum condition for on-column digestion requires the injected trypsin plug centred on the microcolumn. To centre the trypsin plug in the microcolumns, the dead volume of the entire system must be established. This was accomplished by injecting 0.1 M ammonium bicarbonate and monitoring the pH change of the effluent. Dead volumes were established for the entire system in order to optimize the time required for each stage.

Now that proteins have been reduced to peptides and all the excess trypsin and buffers have been removed, the peptides are ready to be eluted from the microcolumn onto a C_8 reverse phase column for separation. If only one microcolumn was coupled to a C_8 separating column, the mass spectrometer would only collect data half the time. During the loading, digestion, and washing steps the mass spectrometer would not be in use. Thus, in order to optimize the automation process, two micro digestion columns, two C_8 columns and three HPLC pumps were required. This enables the mass spectrometer to collect data at all times. This configuration allows for one microcolumn to load, digest and wash while the second microcolumn elutes the peptides from the previous digestion. Once all of these parameters were optimized, the system was capable of running multiple samples without supervision.

4.3.2 On-line Digestion of Proteins

Preliminary investigations of the system focused on manipulation of standard protein samples. The first protein used to characterize the system was cytochrome c. A 100 μL solution of 100 nM cytochrome c was injected onto the R2 microcolumn at 10 μL/min. Once the protein adsorbed onto the reversed-phase beads, a plug of trypsin was injected. The microcolumn, heated to 37 °C, was incubated for 30 min. Following the digestion, peptides were eluted into the mass spectrometer. Figure 4.3 displays the base peak chromatogram of the cytochrome c digest. Each peak in the chromatogram contains several peptides. The most predominate peak gives rise to the intensity used to display the base peak chromatogram. Many of the dominant peaks correspond to peptide fragments of cytochrome c (these peaks are labeled "C"). However, some peaks were from trypsin autolysis (labeled "T" in the figure). MS/MS was performed on the most abundant peptides first; these peptides were then excluded from subsequent MS/MS analysis. This allowed for tandem MS analysis of the lower intensity peptides. The chromatogram displays many cytochrome c peptides separated by the microcolumn and the C_8 separation column. One initial concern with this set-up was excessive band broadening due to the dead volume between the two columns. However, it was found that the peaks are sufficiently resolved for MS/MS and similar separation was achieved when only using one C_8 column. The chromatogram also indicates that little or no undigested cytochrome c was detected. Figure 4.4 displays the fragmentation spectrum

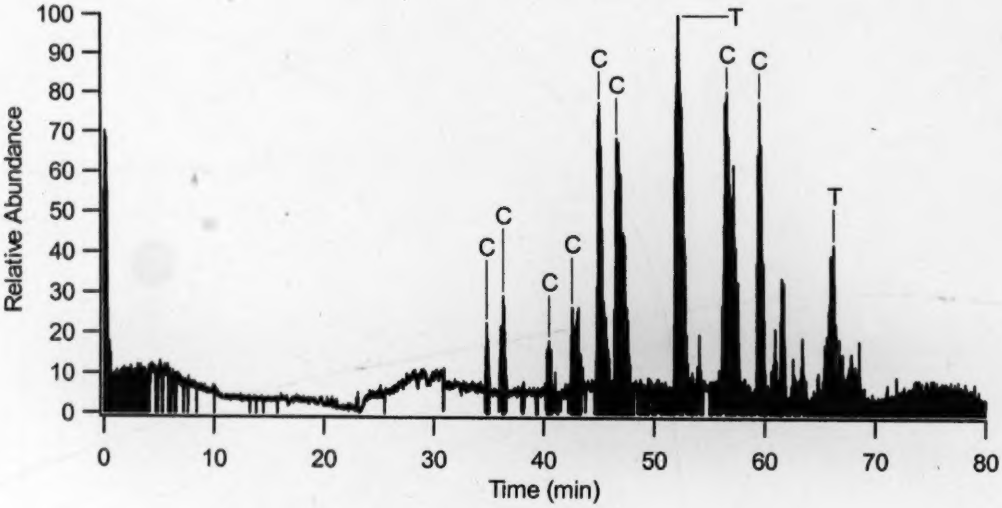


Figure 4.3 Base peak chromatogram of 100 nM cytochrome c digested on-line at 37°C. The peaks labeled "C" represent cytochrome c peptides as the most intense mass spectral peak. Peaks labeled "T" represents trypsin autolysis peptides as the most abundant mass spectral peak.

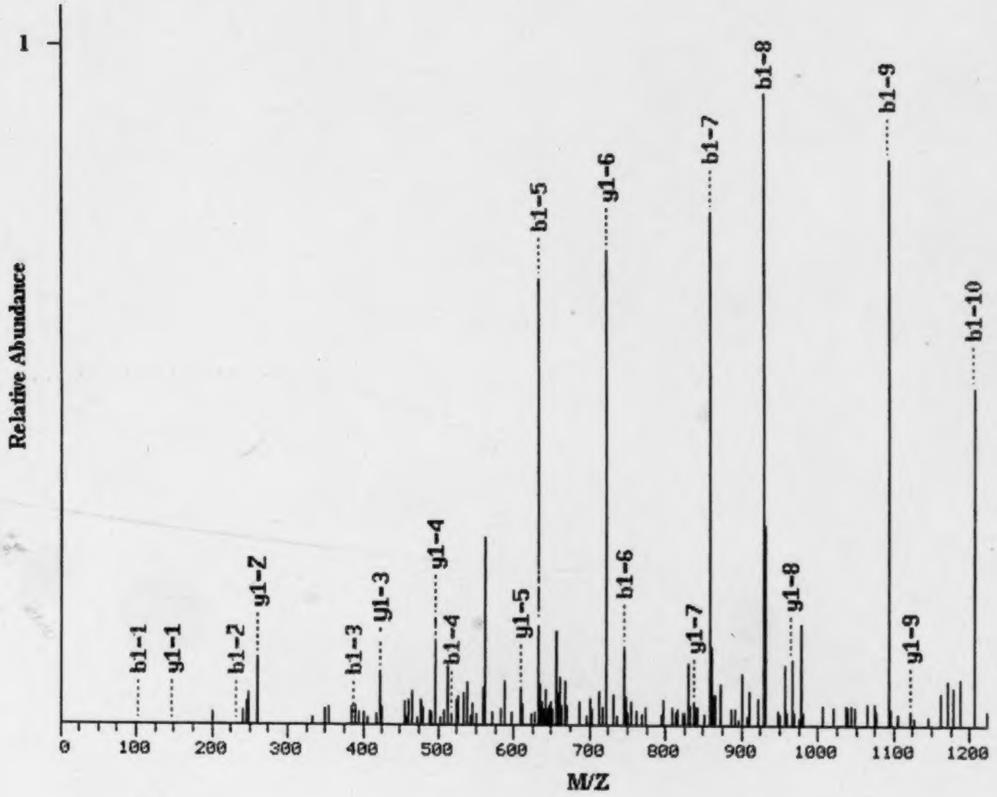


Figure 4.4 Fragmentation pattern of cytochrome c peptide 1352 Da. Ions are labeled with the conventional nomenclature system (Figure 1.7)

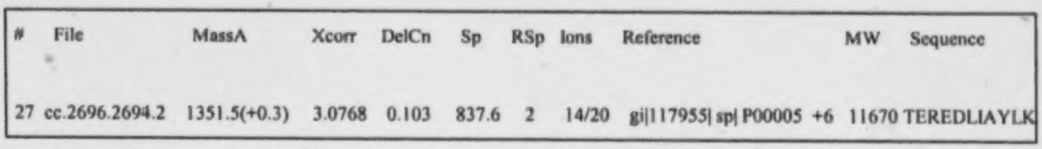
for one of the peptides. The expected y and b-ions are marked. The SEQUEST results for this peptide are displayed in Figure 4.5. Searching the entire chromatogram produced 17 cytochrome c peptides. Each peptide identified had an Xcorr greater than 2. The protein sequence coverage was 82% by amino acid count.

4.3.3 Effect of Temperature on Digestion

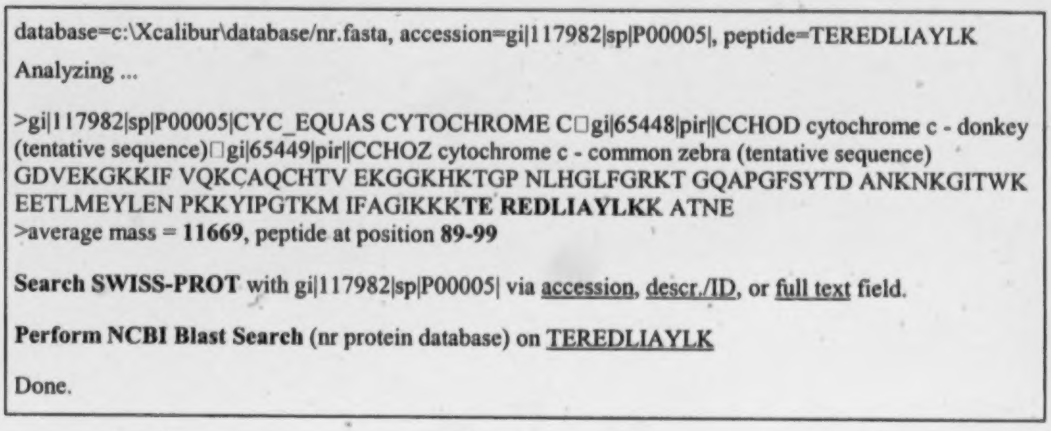
For protein identification, enzymes or chemicals are typically used to fragment proteins into peptides. These peptides are then fragmented further in the mass spectrometer. These segments are in turn detected by the mass spectrometer, thus yielding structural information for the compound. Complete or partial amino acid sequencing of a protein or peptide can be obtained. Thus, it is important that the digestion procedure is efficient. The optimal temperature for trypsin activity is 37°C. In order to achieve an efficient digest, the digestion columns in the automated system were heated. A room temperature digest (Figure 4.6) *vs*. a heated 37°C digest (Figure 4.3) were compared. The two chromatograms are very similar; however the room temperature digest produced 14 cytochrome c peptides, while the 37°C digest yielded 17 peptides. The slight increase in digestion efficiency was expected. Therefore all in-column digests were performed at 37°C.

4.4.4 Re-use of Digestion Column

The intention of this system was to analyze multiple samples in succession. Packing the columns is a simple procedure, but replacing the columns for each run would



#	File	MassA	Xcorr	DelCn	Sp	RSp	Ions	Reference	MW	Sequence
27	cc.2696.2694.2	1351.5(+0.3)	3.0768	0.103	837.6	2	14/20	gi|117955| sp| P00005 +6	11670	TEREDLIAYLK



database=c:\Xcalibur\database/nr.fasta, accession=gi|117982|sp|P00005|, peptide=TEREDLIAYLK
Analyzing ...

>gi|117982|sp|P00005|CYC_EQUAS CYTOCHROME C□gi|65448|pir||CCHOD cytochrome c - donkey (tentative sequence)□gi|65449|pir||CCHOZ cytochrome c - common zebra (tentative sequence)
GDVEKGKKIF VQKCAQCHTV EKGGKHKTGP NLHGLFGRKT GQAPGFSYTD ANKNKGITWK EETLMEYLEN PKKYIPGTKM IFAGIKKKTE **REDLIAYLKK** ATNE
>average mass = **11669**, peptide at position **89-99**

Search SWISS-PROT with gi|117982|sp|P00005| via accession, descr./ID, or full text field.

Perform NCBI Blast Search (nr protein database) on TEREDLIAYLK

Done.

Figure 4.5 SEQUEST search results for cytochrome c peptide with m/z 1352.

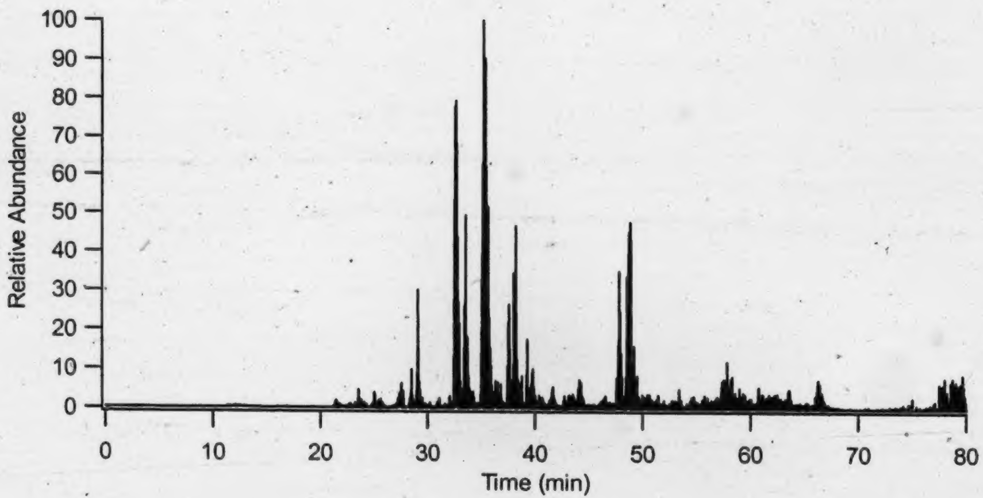


Figure 4.6 Base peak chromatogram of 100 nM cytochrome c digested on-line at room temperature. 14 peptides unique were correlated to cytochrome c.

require time and physical labor. It is impractical to use a new column for each sample; therefore the microcolumns must be re-usable. The major challenge associated with reusing columns is eluting all the proteins and peptides that were previously bound to the column. If there are proteins or peptides left over, they will not only contaminate the next sample but less binding sites would be available for subsequent proteins. Thus, the chromatogram would contain contamination peaks and the peptides of interest would be less intense.

Myoglobin and cytochrome c were used to test the ability to reuse the microcolumns. Myoglobin is a good model protein because it is slightly hydrophobic and relatively difficult to digest with trypsin, compared to cytochrome c. Thus, larger, more hydrophobic peptides would be harder to elute from the column. The column was first loaded with 100 μL of 100 nM cytochrome c. It was digested and eluted from the column, displaying a typical chromatogram similar to Figure 4.3. After the peptides were eluted, the digestion columns were washed and re-equilibrated. A 100-μL solution containing 400 nM myoglobin was injected into the microcolumn. The protein was digested and the base peak chromatogram is displayed in Figure 4.7. A total of 12 myoglobin peptides were identified with no cytochrome c peptides present. The protein sequence coverage was 70% by amino acid count. When comparing the myoglobin and cytochrome c chromatograms, it is apparent that several peptides are eluting later in the myoglobin chromatogram. The dominant peak at 72 min is in fact a myoglobin peptide and the MS/MS spectrum is displayed in Figure 4.8. The Xcorr score for this peptide was 6.4. Following the myoglobin digest, a second cytochrome c digest was performed. This chromatogram was again similar to Figure 4.3, with no myoglobin peptides present.

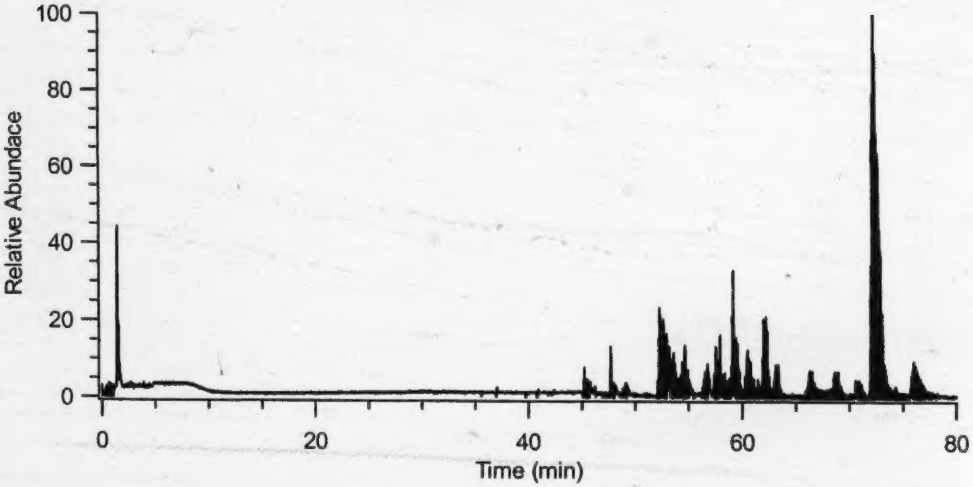


Figure 4.7 Base peak chromatogram of 400 nM myoglobin digested on-line. A total of 12 unique myoglobin peptides were identified.

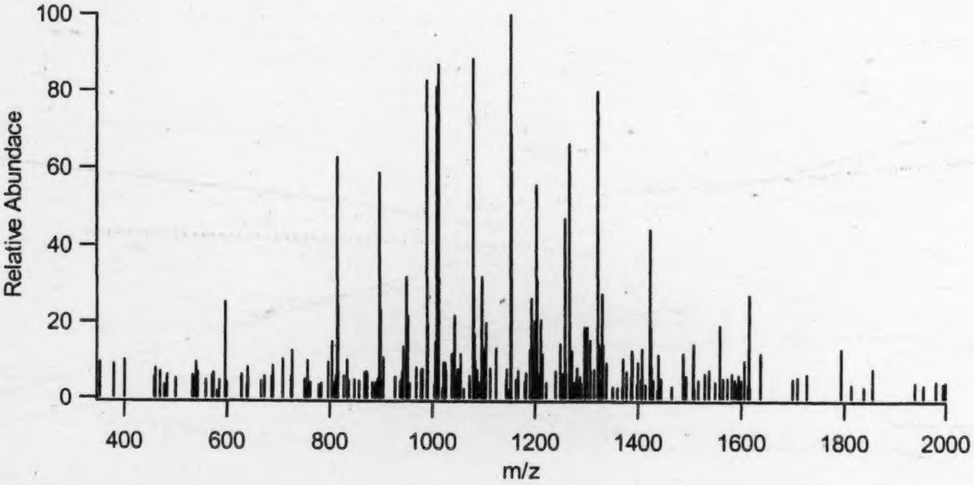


Figure 4.8 Fragmentation of peptide 3405 Da from the myoglobin digest. This peptide eluted from the C_8 separation column at 72 min.

Therefore each microcolumn could be used for multiple samples. Microcolumns were used 30 times without any noticeable deteoriation. However, when dealing with real world samples non-reversible adsorption will eventually affect the column efficiency. Since packing the column was a relatively facile procedure, each microcolumn was replaced after 30-times of usage.

4.4 Conclusions

In previous chapters, a method was developed for the digestion and identification of dilute and contaminated proteins samples. This method displayed encouraging results; however was impractical for multiple sample analysis. The manual labour required to analyze a sample was extensive compared to a typical solution digest. Thus, an automated system was required for this technology to move forward for practical proteomics applications.

The objective of this chapter was to develop an automated online sample manipulation station capable of high throughput analysis. Initially the system required a microcolumn with the following properties: durable, wide pH range tolerance, withstanding HPLC pressures, and coupled easily to HPLC and mass spectrometry. The column created contained all of these characteristics. The in-column digestion efficiency was improved by heating the microcolumns to 37°C.

A system was developed for high throughput analysis. This required the mass spectrometer to collect data continuously. This was achieved through two microcolumns, two C_8 separations columns, and three HPLC pumps. With the use of switching valves, each component worked simultaneously with the rest of the system. The system concentrates and digests proteins, then separates and detects the resulting peptides. Successful digests of cytochrome c and myoglobin were performed in an automated fashion. The on-column digests produced a high number of peptides capable of tandem mass spectrometry. The proteins were easily identified with protein sequence coverage ranging from 70% to 82%. The system successfully ran multiple samples overnight with out supervision. Thus, a system was developed that fully automates protein sample manipulation and is directly coupled with a mass spectrometer.

4.5 Literature Cited

1. Gygi, S. P.; Han, D. K. M.; Gingras, A-C.; Sonenberg, N.; Aebersold, R. *Electrophoresis* **1999**, *20*, 310-319.
2. Lopez, M. F. *J. Chromatogr. B.* **1999**, *722*, 191-202.
3. Klose, J. *Electrophoresis* **1999**, *20*, 643-652.
4. Humphery-Smith, I.; Blackstock, W. *J. Prot. Chem.* **1997**, *16*, 537-544.
5. Anderson, N. G.; Anderson, N. L. *Electrophoresis* **1996**, *17*, 443-453.
6. James, P. *Biochem. Biophys. Res. Commun.* **1997**, *231*, 1-6.
7. O'Farrell, P. H. *J. Biol. Chem.* **1975**, *250*, 4007-4021.
8. Patton, W. F.; Pluskal, M. G.; Skea, W. M.; Buecker, J. L.; Lopez, M. F.; Zimmerman, R.; Belanger, L. M.; Hatch, P. D. *BioTechniques* **1990**, *8*, 518-524.
9. Lopez, M. F.; Patton, W. F. *Electrophoresis* **1997**, *18*, 338-343.
10. Klose, J.; Kobalz, U. *Electrophoresis*, **1995**, *16*, 1034-1059.

11. Nordhoff, E.; Egelhofer, V.; Giavalisco, P.; Eickhoff, H.; Horn, M.; Przewieslik, T.; Theiss, D.; Schneider, U.; Lehrach, H.; Gobom, J. *Electrophoresis* **2001**, *22*, 2844-2855.

12. Garrels, J. I.; McLaughlin, C. S.; Warner, J. R.; Futcher, B.; Latter, G. I.; Kobayashi, R.; Schwender, B.; Volpe, T.; Anderson, D. S.; Mesquita-Fuentes, R.; Payne, W. E. *Electrophoresis* **1997**, *18*, 1347-1360.

13. Hochstrasser, D. F.; Augsburger, V.; Pun, T.; Weber, D.; Pellegrini, C.; Muller, A. F. *Clin. Chem.* **1988**, *31*, 166-170.

14. Nishine, T.; Nakamura, S.; Hazama, M.; Nokihara, K. *Anal. Sci.* **1991**, *7*, 285-288.

15. Nokihara, K.; Morita, N.; Kuriki, T. *Electrophoresis* **1992**, *13*, 701-707.

16. Dhanda, R. K.; Smith, W. M.; Scott, C. B.; Eng, C.; Vijg, *J. Genet. Test.* **1998**, *2*, 67-70.

17. Anderson, N. L.; Anderson, N. G. US Patent 98, **1998**, 59092

18. Bishop, R. US Patent 96, **1996**, 39625

19. Wang, H.; Hanash, S. *J. Chromatogr. B* **2003**, *787*, 11-18.

20. Jeno, P.; Scherer, P. E.; Manning-Krieg, U.; Horst, M. *Anal. Biochem.* **1993**, *215*, 292-298.

21. Pluskal M. G.; Bogdanova, A.; Lopez, M.; Gutierrez, S.; Pitt, A. M. *Proteomics.* **2002**, *2*, 145-150.

22. Porath, *J. Protein Expr. Purif.* **1992**, *3*, 263-281.

23. Hsieh, Y. L.; Wang, H.; Elicone, C.; Mark, J.; Martin, S. A.; Regnier, F. *Anal. Chem.* **1996**, *68*, 455-462.

24. Ericsson, D.; Ekstrom, S.; Nilsson, J.; Bergquist, J.; Marko-Varga, G.; Laurell, T. *Proteomics* **2001**, *1*, 1072-1081.

25. Keller, B.O.; Wang, Z.; Li, L. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2002**, *782*, 317-329.

26. Peterson, D. S.; Rohr, T.; Svec, F.; Frechet, J. M. *Anal. Chem.* **2002**, *74*, 4081-4088.

27. Quadroni, M.; James, P. *Electrophoresis* **1999**, *20*, 664-677.

28. Lopez, M. F.; *Electrophoresis* **2000**, *21*, 1082-1093.

29. Yates III, J.R.; McCormack, A.L.; Schieltz, D.; Carmack, E.; Link, A. *Journal of Protein Chem.* **1997**, *16*, 495-497.

Chapter 5

Automated Online Digestion of *Escherichia coli* Proteins

5.1 Introduction

The goal of proteomics is to map global expression patterns of proteins in cells, tissues or organelles. Since the complete, or nearly completed genome sequences of several organisms have been determined [1], the field of proteomics has expanded exponentially [2]. Libraries or databases created from genomic sequences now allow the entire potential protein complement of organisms to be defined [3]. Mass spectrometric based approaches comprise the most common methods used to identify proteins with high throughput, either through peptide mass mapping or tandem MS of peptides [4-5]. In both of these techniques, the generation of protein fragments (peptides) prior to MS analysis is a necessity, as it is these fragments, rather than the intact protein, that are analyzed by the instrument. The generation of protein fragments is typically accomplished in a controlled fashion with specific enzymes, which cleave the peptide backbone at sequence-defined regions. However, the digestion of complex protein mixtures is problematic. Enzymatic cleavage of proteins is a kinetic process and is therefore influenced by protein concentration; low abundance proteins are cleaved at lower rates than higher abundance proteins [6]. Therefore, in order to overcome this issue, proteins should ideally be separated prior to digestion, allowing for optimized digestion of each individual fraction. Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) is the most powerful protein separation system to date [7]. In recent published reports, 2D-PAGE separations have produced over 1000 resolved spots from *Saccharomyces cerevisiae* on a single gel. However, of these 1000 potentially unique proteins, only approximately 300 were correlated to proteins [8-10]. No reliable information was extracted from the remaining spots. Several reasons contribute to this low level of identification, including dynamic range and proteins solubility issues, which complicate the detection and separation of low-abundance and hydrophobic proteins by 2D-PAGE [11]. Thus, alternative methods for protein separation are being developed [12].

An alternative to 2D-PAGE is multi-dimensional liquid chromatography (LC) [13]. Multi-dimensional HPLC/MS has become a powerful alternative in identifying as many proteins as possible from a complex mixture [14]. This technology applies two complimentary liquid chromatography separation methods in sequence to separate proteins and/or peptides prior to mass spectrometry. One problem with the LC approach is the large number of proteins or peptides that must be separated simultaneously as a prelude to mass spectrometry [15]. Some multi-dimensional systems have reported theoretical peak capacities up to 2500 and separation of up to 1000 proteins bands can be achieved. However, fewer than 50 proteins are usually identified [16-18]. Typically, multi-dimensional LC begins with ion exchange chromatography followed by reversed-phase chromatography. Capillary electrophoresis, gel size exclusion, isoelectric focusing, and affinity chromatography have also been used in the first dimension [13]. In some cases, where the sample is extremely complex, three or more dimensions of separation have been used [19]. However, in the majority of separations, the final stage is reversed-phase chromatography, owing to its compatibility with on-line MS platforms.

There are two different strategies involved with multi-dimensional LC-MS for protein identification. A biological sample can either be digested enzymatically, and the resulting peptides are separated by 2D-LC, or the mixture of proteins can be partially separated in a first dimension prior to digestion, followed by a separation in the second dimension of the resulting peptides generated from the digestion. These methods will be discussed in more detail below.

The first and most common technique involves taking a cell lysate or subcellular fraction and digesting the entire sample prior to separation [20]. The peptide fragments generated from the mixture are separated by ion exchange chromatography followed by reversed-phase chromatography. The advantage of this strategy is the capability of loading high peptide amounts on the ion exchange column. Ion exchange typically has 100 times greater capacity than a reverse phase system, given roughly the same column dimensions [21]. Therefore, higher sample amounts can be loaded onto the ion exchange, allowing ultimately for enhanced detection sensitivity. However, a major disadvantage of this strategy lies in the digestion process. The wide dynamic range of proteins abundances in the whole cell lysate or subcellular fraction makes digestion of the lower abundance proteins difficult. Nevertheless, a micro-scale separation technology employed by Yates *et al.* named MudPIT (Multi-Dimensional Protein Identification Technology) has identified 1,484 proteins from the yeast proteome [14]. The authors analyzed both soluble and insoluble cell lysates resulting in the largest number of identified proteins from a single sample to date.

The second strategy for multidimensional LC-MS involves separation of proteins prior to digestion. Typically, this approach employs ion exchange chromatography as the first dimension of protein separation. These fractions can then be digested, or alternatively, undergo further separation (with reversed-phase chromatography for example) prior to protein digestion. This approach aids in the dynamic range issues associated with the digest. However, the method does result in a significant increase in the amount of time required for sample processing, as one must now digest each individual fraction collected from the LC separation. Hu *et al.* have applied this strategy to identify proteins in *E. coli* [22]. The authors extracted proteins from *E. coli* under different conditions and combined the results to identify 310 unique proteins. Recently, Hunt *et al.* employed a method similar to MudPIT and identified 986 soluble proteins from the extraction of *E. coli* [23]. This is the largest set of protein identifications from *E. coli* to date.

Escherichia coli is an important model organism for biochemical and biological studies as its entire genome has been sequenced since 1997 [24] and is fully available in genome databases such as SWISS-PROT [25]. The *E. coli* proteome is considerably smaller than that of yeast; nonetheless, it has been revealed to date that at least 1000 different proteins are present in *E. coli* [26]. However, it has been estimated that *E. coli* can theoretically express more than 4000 proteins, based on the genetic sequence of the organism [27]. The analysis of a sample of this complexity requires the need for a robust, automated, and sensitive high-throughput technology. The research goal of this chapter is to develop an automated technique to identify proteins, with high throughput capabilities.

In Chapter 4, an automated approach was described for the capture of proteins in a dilute solution onto a reversed-phase column. The bound proteins were subjected to trypsin proteolysis and the peptides were eluted from the column for identification by mass spectrometry in a potentially high-throughput fashion. In the work presented herein, the application of this technique to high-throughput proteomic analysis is illustrated for the identification of proteins from *E. coli* whole cell extracts.

5.2 Experimental

5.2.1 Materials and Reagents

Trypsin (from bovine pancreas, TPCK treated to reduce chymotrypsin activity; dialyzed, lyophilized) was from Sigma Aldrich Canada (Oakville, ON). Poros R2 beads (20 μm) were from PerSeptive Biosystems (a gift of Professor F. Cantwell, University of Alberta). All PEEK tubing, unions and frits were purchased from Upchurch (Oak Harbor, WA, USA). Analytical grade acetone, methanol, acetonitrile, acetic acid, and trifluoroacetic acid (TFA) were purchased from Caledon Laboratories (Edmonton, AB,). Water used in the experiments was from a NANOpure water system (Barnstead/Thermolyne). All other chemicals were purchased from Sigma Aldrich Canada.

5.2.2 Growth of *E. coli*

E. coli (ATCC 47076) was purchased from American Type Culture Collection (ATCC). The starter culture was prepared by inoculating a 50 mL started culture comprising 50 mL of nutrient broth. The composition of nutrient broth was: beef extract 3.0 g/L and peptone 5.0 g/L. The culture was grown at 37°C for 2 hours with agitation.

This starter culture was added into 1550 mL of LB media. The cells were harvested once an OD of 6 was reached (48 hours). The cells were isolated by centrifugation at 15344 × *g* for 20 minutes and were subsequently washed with Milli-Q water once, lyophilized in a lyophilizer and stored at -20°C until ready for analysis.

5.2.3 Cell Lysate

Seven 6-mg portions of lyophilized *E. coli* cells were weighed into 7 separate vials, and the proteins were extracted with 2 mL of 100 mM Tris/HCl buffered at pH 7.8. For each vial, the cells were ruptured by sonicating for one minute in pulsed mode at 1 pulse per second with pulse durations of 0.75 seconds while on ice (Branson Sonifier, Danbury, CT, USA). The cells were then spun down at 14,000 × *g* for 20 min. The supernatant was filtered through a Microcon-3 filter (Millipore), with a molecular weight cutoff of 3000 Da, and the portion of the sample which did not pass through the filter was collected in a final volume of 70 μL. The cell lysates from each Microcon filtration were combined, resulting in a total retrieved volume of ~ 500 μL. A Bradford protein assay was used to quantify the amount of protein present in this fraction. The concentration was determined to be 40 μg/μL. Therefore, 20 mg of protein was extracted from 42 mg of lyophilized cell pellet. The protein in the fraction was then subjected to treatment with 1.2 mg of dithiothreitol (DTT) at 56°C for 1 hour to reduce the disulfide bonds. The free cysteines in the proteins were then alkylated with 2.4 mg of iodoacetamide at room temperature for 1 hour.

5.2.4 Separations

5.2.4.1 Anion Exchange Chromatography

The reduced and alkylated protein mixture was separated using a 4.6 × 50 mm Mono Q HR 5/5 column from Amersham Pharmacia Biotech (on loan from IBD, University of Alberta). The anion exchange beads were 10 μm in diameter with a column capacity of 50 mg. The entire 500 μL aliquot, comprising 20 mg of *E. coli* protein, was injected into the column. A flow rate of 1 mL/min was used and fractions were collected in one-minute increments. A gradient separation was performed using water with 20 mM Tris/HCl at pH 8 (solvent A), or with the addition of 1 M NaCl (solvent B). The gradient used was as follows: hold B at 0% for 5 min, increase B from 0 to 35% over 30 min, then from 35 to 50% B over 15 min, then increase B to 100% over 20 min. A total of 50 fractions were collected, and a Bradford protein assay was used to quantify the amount of protein present in each collected fraction. The fractions were then placed in the HPLC auto-sampler for further analysis.

5.2.4.2 Reversed-Phase Chromatography

Following digestion of the proteins (see section 5.2.6) a Vydac reversed-phase n-octylsilica column (Anaheim, CA, USA), with dimensions of 150 μm × 15 cm was used to separate peptides in the second dimension. A gradient separation was performed using water with 0.5% acetic acid (solvent A) and acetonitrile with 0.5% acetic acid (solvent B). The gradient used was as follows: increase B from 5 to 10% over 2 min, then from 10 to 45% B over 38 min, then increase B to 65% over 10 min.

5.2.5 Sample Manipulation on Microcolumn

The proteins in the cell lysate were first separated by anion exchange chromatography. A Bradford protein assay was used to quantify the amount of protein present in each fraction. The injection volume of each fraction was calculated to inject 1 μg of protein. This amount was then loaded onto the microcolumn for clean up and concentration. The procedure for sample manipulation is previously described in Chapter 4. Briefly, following a washing step, buffered trypsin was introduced into the microcolumn. The adsorbed proteins were digested for 30 min at 37°C. The microcolumn was then washed prior to eluting the peptides onto a C_8 analytical column.

5.2.6 Solution Digestion

To compare solution conditions, fraction 29 min of the anion exchange separation was digested in solution. A 20 μL aliquot was digested with 0.5 μg of trypsin. The solution was allowed to incubate at 37°C for 24 hours. Following the completion of the digest, a ZipTip (Millipore, Burlington, MA, USA) was used to clean and concentrate the sample. A 2 μL aliquot of 50% (v/v) acetonitrile/ 0.1% TFA (v/v) was used to elute the peptides from the ZipTip. The eluent was then injected into a C_8 column for separation and a LCQ^{Deca} was used for detection.

5.2.7 Instrumentation

An Agilent 1100 (Palo Alto, CA, USA) chromatographic system was used for the anion exchange chromatography. UV chromatograms were recorded at 210 nm. The automated sample manipulation system was illustrated in Chapter 4 (Figure 4.2). *E. coli* proteins were loaded onto the microcolumns, as previously described in Chapter 4 (Figure 4.1), with the autosampler of the HP 1100 system. A second Agilent 1100 pump was used to wash one of two Vydac n-octylsilica (C_8) columns (Anaheim, CA, USA). A third pump, Thermo Finnigan Surveyor (San Jose CA, USA), was used to elute peptides from the micro and C_8 columns into the mass spectrometer. Switching valves used in the automated system to divert flow were from Rheodyne (Rohnert Park, CA, USA). A Finnigan LCQ^{Deca} ion trap mass spectrometer (San Jose, CA, USA) was used to detect and perform MS/MS on the peptides. Data analysis was performed by SEQUEST, as described in Chapter 4. The MS/MS data was searched against the ecoli.fasta database from the National Centre for Biotechnology. This database was indexed for speed with 3 missed cleavages of trypsin. The criteria for protein identification was Xcorr scores higher than 2.0 [27].

5.3 Results and Discussion

5.3.1 Pre-Fractionation Prior to Digestion

Proteins were extracted from the *E. coli* cells using tris/HCl buffer at pH 7.8. The *E. coli* lysate contains a complicated mixture of proteins. The cell lysate was first analyzed with a one-dimensional gel to verify that the extraction was successful. The extracted proteins were separated by a 1-D gel and stained with commassie blue. The 1D gel is displayed in Figure 5.1. From the gel, it is apparent that many proteins were

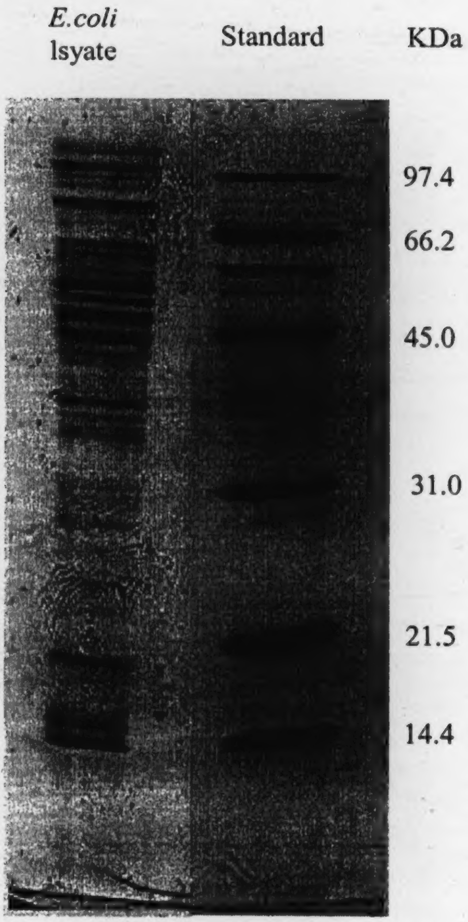

E.coli
lsyate
Standard
KDa
97.4
66.2
45.0
31.0
21.5
14.4

Figure 5.1 SDS PAGE image of proteins extracted from *E. coli.* with 100 mM tris/HCl buffered at pH 7.8.

successfully extracted from *E. coli*. As previously discussed, complex biological samples require separation prior to digestion. Therefore, anion exchange chromatography (AEC) was used as an initial separation of the proteins prior to digestion. The buffer conditions for AEC were set to match the conditions of the extraction buffers to prevent precipitation of proteins. The UV absorbance chromatogram obtained from the separation of the *E. coli* protein mixture is displayed in Figure 5.2. The chromatogram displays the large number of proteins extracted from *E. coli*. Judging from the number of peaks in the chromatogram, anion exchange chromatography has resulted in a partial, yet effective, separation of the protein mixture.

5.3.2 Digestion of Anion Exchange Fractions

From the UV chromatogram (Figure 5.2), proteins are present in fractions 3 to 50 min. As discussed in previous chapters, the importance of loading the correct amount of protein onto a microcolumn for digestion is paramount. Thus, each fraction was quantified prior to sample loading. The UV absorbance chromatogram could be used to estimate how much protein is present in each fraction; however, a Bradford protein assay would give a more accurate measure of the protein concentration. The results from the protein assay are displayed in Table 5.1. From this table, the total amount of protein per fraction ranged from less than 5 – 60 μg/mL. In previous chapters 0.12 μg of cytochrome c was used as a standard, however each *E. coli* fraction is expected to contain numerous proteins. The concentration of an individual protein in the fraction would therefore be relatively low in comparison. To compensate for this, 4 μg of total protein from each fraction was injected into the microcolumn for digestion. A total of 43 fractions

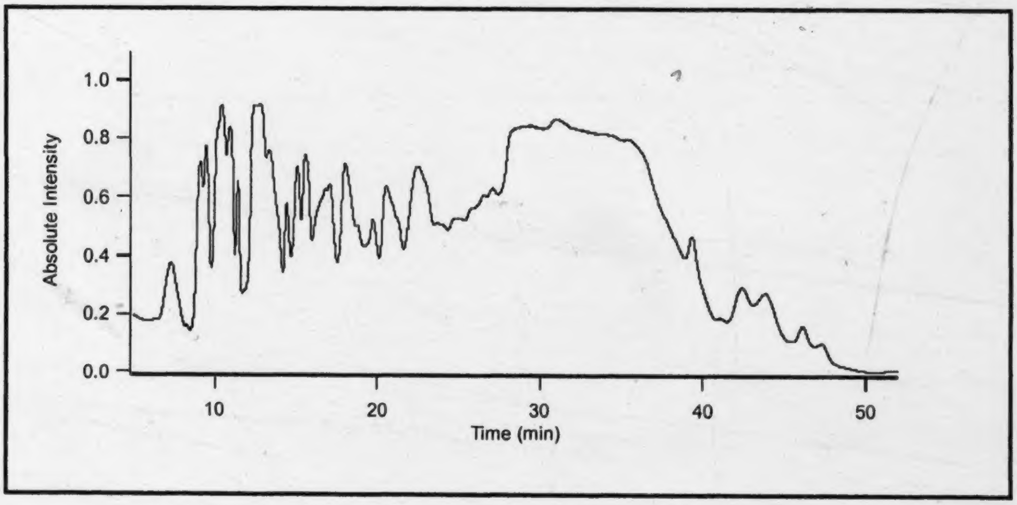


Figure 5.2 UV chromatogram of *E. coli* lysate separated on a Mono Q HR 5/5 column. Fractions 8 – 50 were subsequently digested for protein identification.

Table 5.1 Bradford protein assay results from anion exchange fractions

Fraction (min)	Concentration µg / mL	Volume Injected µL
8	<5	100
9	<5	100
10	<5	100
11	45	100
12	<5	100
13	240	20
14	320	15
15	180	20
16	200	20
17	85	50
18	200	20
19	150	35
20	150	35
21	220	20
22	510	20
23	320	10
24	230	20
25	261	20
26	350	15
27	320	15
28	250	20
29	270	20
30	520	10
31	600	10
32	460	10
33	190	20
34	210	20
35	210	20
36	160	25
37	91	50
38	86	50
39	85	50
40	5	100
41	14	100
42	82	50
43	74	50
44	30	75
45	<5	100
46	<5	100
47	<5	100
48	<5	100
49	<5	100
50	<5	100

required analysis. The system is capable of analyzing one sample every 70 min; therefore a total of 58 hours was required to process all fractions. Each fraction was loaded into the system as described in Chapter 4. The resulting peptides were then eluted from the digestion column onto a C_8 reverse phase column for separation and detected by a LCQ^{Deca}.

5.3.3 Analysis

Figure 5.3 displays the base peak chromatograms from the LC-ESI analysis of several digested *E. coli* fractions. The figure illustrates that online column digests produced many peptides. Fractions 9, 22, 29, and 44 minutes were selected as representative fractions and are displayed in Figure 5.3. The SEQUEST algorithm was used to interpret the MS/MS data. Since the species, *E. coli*, is known, the ecoli.fasta database can be used to reduce the searching time. SEQUEST requires 30 min to analyze each chromatogram. Fractions 9, 22, 29, and 44 minutes identified 4, 4, 9, and 5 proteins respectively. Figure 5.4 displays a typical SEQUEST output resulting from searching the *E. coli* database. An Xcorr value of 2.0 or greater was used to identify proteins. This is consistent with the search criteria described previously by Yates *et al.* [27]

The majority of proteins identified were found to appear over several fractions from the first dimension anion exchange separation. An average of 5 proteins were identified per fraction. A total of 199 proteins were identified from the 43 fractions, however, only 68 were unique. Upon initial comparison, this amount appears to be low, relative to other techniques. Hu *et al.* identified 310 unique proteins using two-dimensional separation prior to digestion; however, this total was obtained from

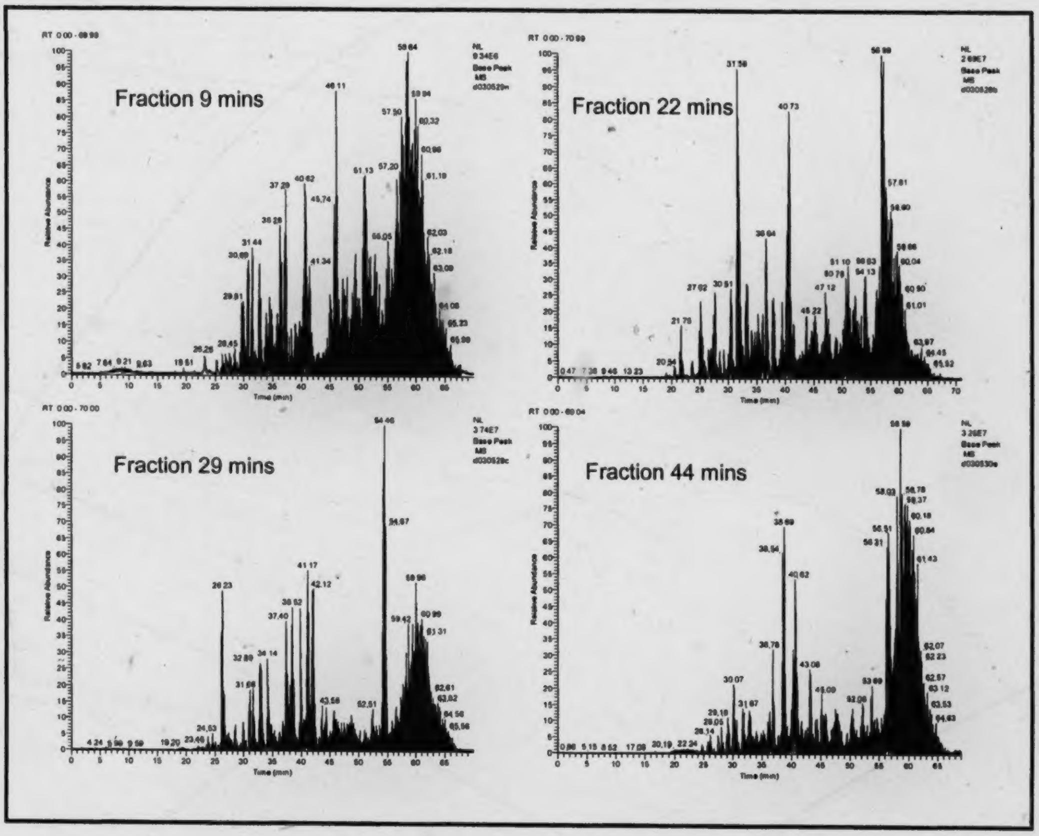
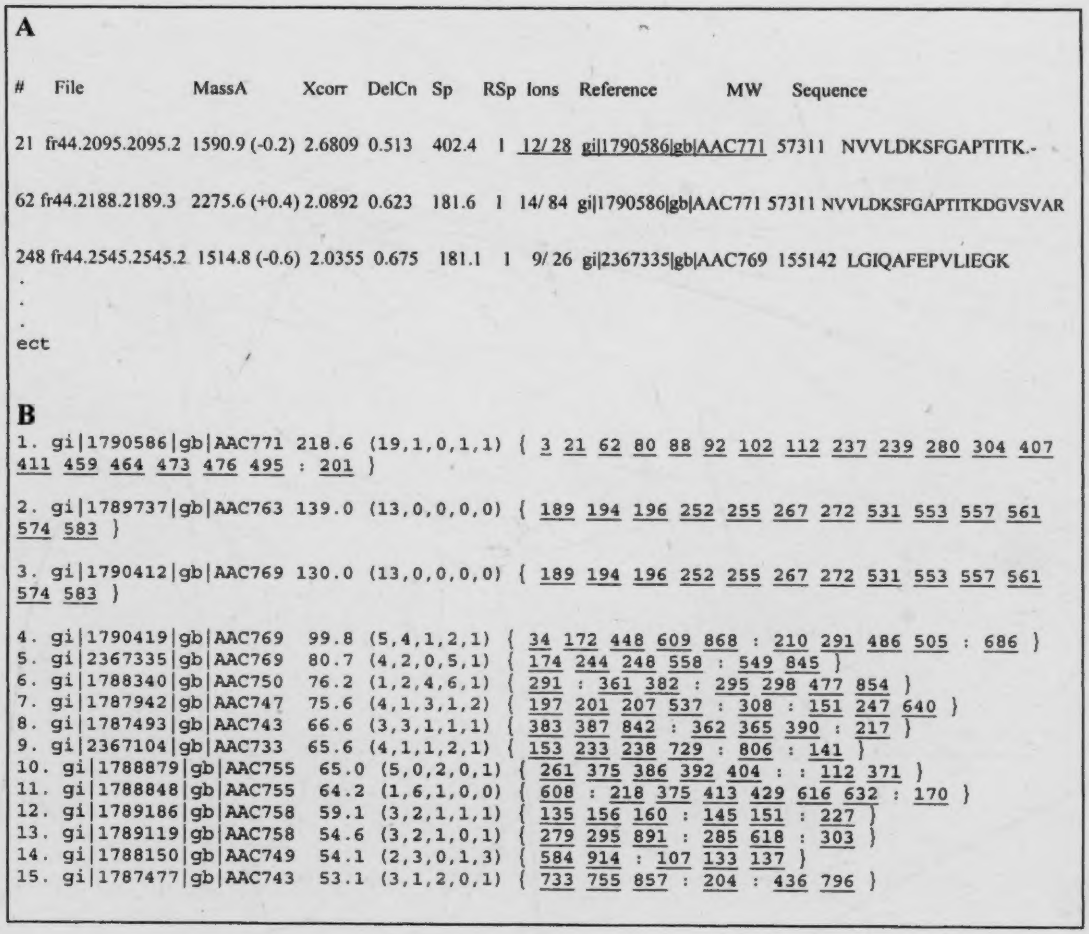


Figure 5.3 Base peak chromatograms from fractions 9, 22, 29, and 44. A total of 4, 4, 9, 5 proteins were identified in each respective fraction.



A

```
#  File              MassA          Xcorr  DelCn  Sp    RSp Ions   Reference            MW     Sequence

21 fr44.2095.2095.2  1590.9 (-0.2)  2.6809 0.513  402.4  1  12/ 28 gi|1790586|gb|AAC771 57311  NVVLDKSFGAPTITK.-

62 fr44.2188.2189.3  2275.6 (+0.4)  2.0892 0.623  181.6  1  14/ 84 gi|1790586|gb|AAC771 57311  NVVLDKSFGAPTITKDGVSVAR

248 fr44.2545.2545.2 1514.8 (-0.6)  2.0355 0.675  181.1  1   9/ 26 gi|2367335|gb|AAC769 155142 LGIQAFEPVLIEGK
.
.
.
ect
```

B

```
1. gi|1790586|gb|AAC771 218.6 (19,1,0,1,1) { 3 21 62 80 88 92 102 112 237 239 280 304 407
411 459 464 473 476 495 : 201 }

2. gi|1789737|gb|AAC763 139.0 (13,0,0,0,0) { 189 194 196 252 255 267 272 531 553 557 561
574 583 }

3. gi|1790412|gb|AAC769 130.0 (13,0,0,0,0) { 189 194 196 252 255 267 272 531 553 557 561
574 583 }

4. gi|1790419|gb|AAC769  99.8 (5,4,1,2,1) { 34 172 448 609 868 : 210 291 486 505 : 686 }
5. gi|2367335|gb|AAC769  80.7 (4,2,0,5,1) { 174 244 248 558 : 549 845 }
6. gi|1788340|gb|AAC750  76.2 (1,2,4,6,1) { 291 : 361 382 : 295 298 477 854 }
7. gi|1787942|gb|AAC747  75.6 (4,1,3,1,2) { 197 201 207 537 : 308 : 151 247 640 }
8. gi|1787493|gb|AAC743  66.6 (3,3,1,1,1) { 383 387 842 : 362 365 390 : 217 }
9. gi|2367104|gb|AAC733  65.6 (4,1,1,2,1) { 153 233 238 729 : 806 : 141 }
10. gi|1788879|gb|AAC755  65.0 (5,0,2,0,1) { 261 375 386 392 404 : : 112 371 }
11. gi|1788848|gb|AAC755  64.2 (1,6,1,0,0) { 608 : 218 375 413 429 616 632 : 170 }
12. gi|1789186|gb|AAC758  59.1 (3,2,1,1,1) { 135 156 160 : 145 151 : 227 }
13. gi|1789119|gb|AAC758  54.6 (3,2,1,0,1) { 279 295 891 : 285 618 : 303 }
14. gi|1788150|gb|AAC749  54.1 (2,3,0,1,3) { 584 914 : 107 133 137 }
15. gi|1787477|gb|AAC743  53.1 (3,1,2,0,1) { 733 755 857 : 204 : 436 796 }
```

Figure 5.4 Typical SEQUEST search results for anion exchange fraction digestion on beads in the automated system. Section A lists every MS/MS cycle performed. Section B summarizes the data from section A.

extraction of *E. coli* under several different conditions [22]. Also, 2-Dimensional HPLC was performed prior to digestion. Recently, Hunt *et al.* identified 986 proteins form *E. coli* whole cell extract using technology similar to MudPIT [23].

However, these are not direct comparisons to my method. In our lab, *E. coli* has recently been analysed by 2-D LC MS/MS [28]. In these results, a range of 88-160 proteins were identified depending on the solution digestion conditions. This is a direct comparison, as the same *E. coli* cells were used with the same extraction procedure. Clearly the source of the sample and the method of extraction are important in determining the actual number of proteins that can be identified by MS. Another possibility for the low number of proteins identified, relative to the total number of proteins present in *E. coli* is the limited detection dynamic range of the system due to insufficient separation of proteins prior to automated digestion and MS/MS analysis.

The majority of proteins identified were present over several fractions. Thus each fraction contains many proteins. Knowing high abundance proteins are digested at a faster rate then low abundance proteins; the mass spectra will be dominated by the peptides resulting from the high abundance proteins.

5.3.4 Method Comparison

This system provides a means to concentrate, clean, and digest protein samples in a fully automated fashion. With this set-up (Figure 4.2) the system was capable of analyzing 43 sequential anion exchange fractions in 58 hours. The resulting digests produced peptides that were fragmented and the MS/MS data was used for protein identification. This method was compared to an off-line solution digest of an anion exchange fraction followed by sample concentration with a ZipTip. Fraction 29 was used for comparison. The same amount of total protein was used for the solution digest (1 μg). Digesting, this amount, for 2 hrs at 37 °C resulted in no peptides being detected by the LCQ$^{\mathrm{Deca}}$. Since the concentration is low, the solution digest conditions require 24 hrs to complete. Once the digestion was completed the sample was cleaned and concentrated with a ZipTip. A 2 μL aliquot was used to elute the sample from the ZipTip. The aliquot was manually injected onto a C_8 column for separation and detected with an LCQ$^{\mathrm{Deca}}$. Figure 5.5 displays the base peak chromatogram resulting from the digest. SEQUEST search results identified 9 proteins. Both on-column and solution digests identified 9 proteins; however, only 5 of the 9 proteins are identical. This can be expected, since each protein will digest differently on the reversed-phase surface *vs.* in solution. The differences between solution digestion and digestion on a surface were thoroughly discussed in Chapter 3. The on-column digestion method gives a slightly different picture thus providing complementary protein identification information to the traditional solution digestion method.

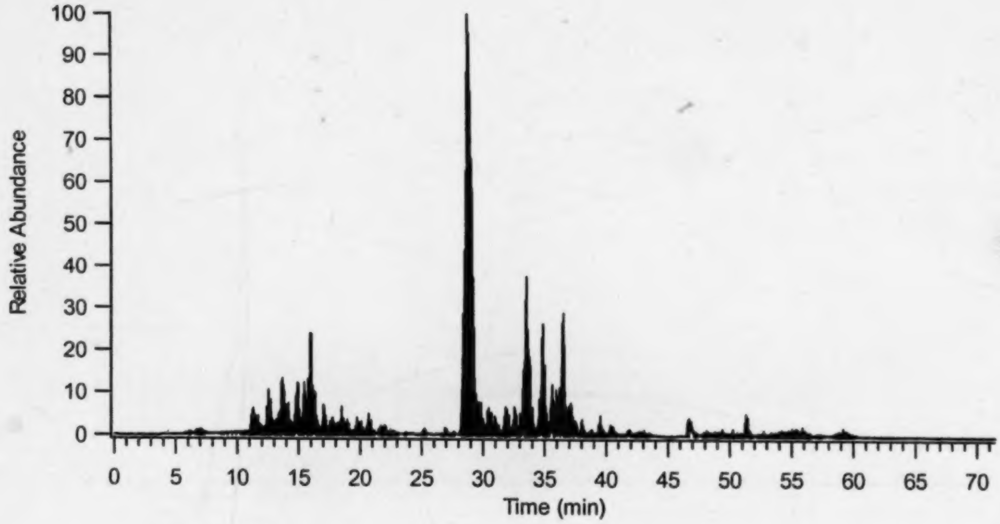


Figure 5.5 Base peak chromatogram of fraction 29 minutes, from anion exchange separation. Digestion was performed in solution and concentrated with ZipTip prior to injection. A total of 9 proteins were identified when searched against the ecoli.fasta database.

5.4 Conclusions

Proteins were successfully extracted from *E. coli*. The whole cell lysate was separated on an anion exchange column. The fractions were collected and quantified by a Bradford protein assay. Once the amount in each fraction was known, the appropriate amount was loaded onto microcolumn for digestion. Digests were achieved in the microcolumn and detected by ESI in a LCQ$^{\mathrm{Deca}}$ ion trap in an automated fashion. Total time for analysis was 58 hours for 43 samples. The tandem MS data collected from each sample was interpreted with the SEQUEST algorithm. Numerous proteins were identified in each fraction. Typically, peptides used for the identification purposes were seen over several fractions. A total of 68 unique proteins were identified with this method. One possibility that only 68 proteins were identified was the dynamic range of proteins in each fraction. Since the majority of peptides were seen over several anion exchange fractions, greater separation in the first dimension is required. To increase the total number of peptides identified, perhaps two-dimensional HPLC is required prior to on-column digestion. Also, only proteins soluble in a basic aqueous buffer (100 mM Tris, pH 8) were analyzed. If the goal is to analyze the entire proteome, multiple extractions using different conditions should be performed. This would include acidic conditions as well as surfactants for insoluble proteins, such as membrane proteins. Hu *et al.* used several extraction methods combined with two-dimensional HPLC prior to digestion to identify 310 proteins. The strategy of multiple extractions and 2D HPLC prior to digestion is a very labour-intensive undertaking. However, this method demonstrates it has the capability of high throughput analysis, enabling such labour extensive strategies. Thus current effort will be focused on 2-Dimensional separation prior to digestion or improving separation efficiency in one-dimensional separation of protein. In addition, increasing the dynamic range of on-column digestion is another key to improve the overall performance of the automated system for proteome analysis.

5.5 Literature Cited

1. Devine, K. M.; Wolfe, K. *Trends Genet.* **1995**, *11*, 429-431.
2. Phizichy, E.; Bastiaens, P. I. H; Zhu, H.; Snyder, M.; Fields, S. *Nature* **2003**, *442*, 208- 215.
3. James, P.; Quadroni, M.; Carafoli, E.; Gonnet, G.; *Protein Science* **1994**, *3*, 1347-1350.
4. Henzel, W. J.; Billeci, T. M.; Stults, J. T.; Wong, S. C.; Grimley, C.; Watanabe, C. *Proc. Natl. Acad. Sci.* USA **1993**, *90*, 5011-5015.
5. Rosche, F.; Schmidt, J.; Hoffmann, T.; Pauly, R.P.; McIntosh, C. H.; Pederson, R.A.; Demuth, H. U. *Methods Mol Biol.* **2000**, *146*, 251-272.
6. Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3351-3367.
7. O'Farrell, P. H. *J. Biol. Chem.* **1975**, 250, 4007.
8. Perrot, M.; Sagliocco, F.; Mini, T.; Monribot, C.; Schneider, U.; Shevchenko, A.; Mann, M.; Jeno, P.; Boucherie, H. *Electrophoresis* **1999**, *20*, 2280-2298.

9. Futcher, B.; Latter G. I.; Monardo P.; McLaughlin, C. S.; Garrels, J. I.; *Mol. Cell. Biol.* **1999**, *19*, 7357-7368.

10. Garrels, J. I.; McLaughlin, C. S.; Warner, J. R.; Futcher, B.; Latter, G. I.; Kobayashi, R.; Schwender, B.; Volpe, T.; Anderson, D. S.; Mesquita-Fuentes, R.; Payne, W. E. *Electrophoresis* **1997**, *18*, 1347-1360.

11. Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C. *Electrophoresis* **2001**, *21*, 1104.

12. Patterson, S. D.; *Curr. Opin. Biotechnol.* **2000**, 11, 413-418.

13. Wang, H.; Hanash, S.; *J. Chromatogr. B* **2003**, *787*, 11-18.

14. Washburn, M.P.; Wolters, D.; Yates III, J.R. *Nat. Biotechnol.* **2001**, 19, 242–247.

15. Davis, J. M.; Giddings, J.C. *Anal. Chem.* **1983**, *55*, 418-424.

16. Opitek, G. J.; Lewis, K. C.; Jorgenson, J. W.; Anderegg, R. J. *Anal. Chem.* **1997**, *69*, 1518-1524.

17. Wall, D. B.; Kachman, M. T.; Gong, S.; Hinderer, R.; Parus, S.; Misek, D. E.; Hanash, S. M.; Lubman, D. M. *Anal. Chem.* **2000**, 72, 1099-1111.

18. Washburn, M. P.; Yates III, J. R. *Proteomics. A Trends Guide* (Mann, M., Blackstock, W., Eds.) **2000**, 28-32.

19. Wagner, K.; Miliotis, T.; Marko-Varga, G.; Bischoff, R.; Unger, K.K; *Anal. Chem.* **2002**, *74*, 809-820.

20. Link, A. J.; Eng, J.; Schieltz, D. M.; Carmack, E.; Mize, G. J.; Morris, D. R.; Garvik, B. M.; Yates III, J. R. *Nat. Biotechnol.* **1999**, *17*, 676-682.

21. Liu, H.; Lin, D.; Yates III, J. R. *Biotechniques* **2002**, *32*, 898.

22. Champion, M. M.; Campbell, C. S.; Siegele, D. A.; Russell, D. H.; Hu, J. C. *Molecular Microbiology* **2003,** *47*, 383-396.

23. Corbin, R. W.; Paliy, O.; Yang, F.; Shabanowitz, J.; Platt, M.; Lyons, C. E. Jr.; Root, K.; McAuliffe, J.; Jordan, M. I.; Kustu, S.; Soupene, E.; Hunt, D. F. *Proc. Natl. Acad. Sci. U S A* **2003**, *100*, 9232-9237.

24. Blattner, F. R.;Plunkett, G.; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. *Science* **1997**, *277*, 1453-1474.

25. Bairoch A., Apweiler R., *Nucleic Acids Res.* **2000**, *28*, 45–48.

26. P. O'Farell, *J. Biol. Chem.* **1975**, *250*, 4007.

27. Pennington, S.R.; Wilkins, M. R.; Hochstrasser, D. F.; Dunn, M. J.; *Trends in Cell Biology* **1997**, *7*, 168.

28. Yates III, J.R.; McCormack, A.L.; Schieltz, D.; Carmack, E.; Link, A. *Journal of Protein Chem.* **1997**, *16*, 495-497.

29. Chen, R.; Tao, L.; Zhang, H.; Wishart, D.; Li, L. *ASMS Proceedings* **2003,** poster 449.

Chapter 6

Conclusions and Future Work

The work embodied in this thesis represents further development of surface digestion technology into an automated platform for global protein identification by mass spectrometry. In previous work [1], the surface digestion method was applied to MALDI-MS. This method essentially involved absorbing proteins to a hydrophobic support, in the form of reversed-phase chromatographic packing material (micro beads). Following protein adsorption, the beads were washed to remove various contaminants, and the appropriate enzyme was added to the bead slurry in order to induce digestion of the surface-bound proteins. Peptide mass mapping experiments were conducted by placing the beads directly onto a MALDI target for analysis. Although this method was ideally suited for protein identification experiments by mass spectrometry, the procedure was neither directly amenable to ESI mass spectrometry, nor to high throughput protein characterization. This thesis has focused on extending the surface digestion technology described above to ESI applications and to the development of an automated on-line sample manipulation system.

In Chapter 2, the principle of surface digestion for protein characterization was extended to a column format, thus allowing for ESI analysis of the protein samples. In this method, the reversed-phase chromatographic material was packed inside a thin pipette tip in order to form a microcolumn of the hydrophobic support. With this device, a procedure was developed for protein identification using mass spectrometry that incorporates sample cleanup, pre-concentration, and protein digestion in a single-stage system. Sample loading was accomplished by flowing the protein solution through the microcolumn, where the protein adsorbs to the hydrophobic surface. The protein was digested while still bound to the hydrophobic support by flowing a buffered enzyme solution through the column bed. The peptide fragments were subsequently eluted for detection by MALDI, or by ESI-MS. The procedure was applied to the MALDI-MS/MS identification of proteins present in an individual fraction collected by ion exchange HPLC separation of *E. coli* total cell extract. An additional application was illustrated with the analysis of a human plasma fraction. A total of 14 proteins, which were present in the sample at sub-micromolar concentrations, were identified from ESI-MS/MS. Also human tears were analyzed with this system and 4 proteins were detected. This chapter not only illustrated the feasibility of surface bound protein digestion in a microcolumn format, but also demonstrated that the procedure could be applied to complex biological matrices, such as the human plasma sample.

In Chapter 3, an in-depth study was performed on the effects of surface chemistry and morphology as pertaining to the digestion of surface bound proteins. Here the peptides generated from digestion of proteins bound to various types of micro-bead surfaces were examined in order to determine if the type of surface will influence the digestion process. Detailed examinations of site cleavages and sequence coverage were carried out for a tryptic digestion of cytochrome c adsorbed on reversed-phase polystyrene divinylbenzene (Poros R2 beads) versus C_{18} bonded-phase silica beads. It was shown that although the surface did not completely hinder the digestion of cleavage sites of the protein, the digestion products were clearly different than those obtained from a solution digest. Specifically, partial digestion was obtained from surface digestion, resulting in a greater number of missed cleavages than a comparable solution digest. Subsequent comparisons of peptide mass maps generated from the digestion of various proteins on surfaces with altering chemistry (C_4, C_8, C_{18}, and R2 beads), or with different surface morphology, were performed. The results reveal that surface chemistry plays only a minor role in affecting the peptide mass maps, and surface morphology has no noticeable effects on the resulting peptide mass maps. It was also shown that the mass spectrometric detection method used to analyze the digested peptides could significantly influence the information content on cleavage sites and the extent of sequence coverage. The use of a combination of MALDI, LC/off-line MALDI, and LC/ESI MS was demonstrated to be crucial in revealing subtle changes in the peptide mass maps. Surface digestion was also extended to the use of various other digestion reagents, other than trypsin. The enzymes chymotrypsin, pepsin, and the chemical cleavage reagent, cyanogen bromide (CNBr) were successfully used to fragment adsorbed proteins. The resulting peptide mass maps were similar to digestion performed in solution. In addition, these digestion approached could be applied in tandem, as was demonstrated from a combined trypsin/CNBr digestion of bacteriorhodopsin.

In Chapter 4, the technology described above for in-column digestion of proteins was further developed into an automated system capable of high throughput analysis. This was accomplished by constructing a reusable column capable of easy coupling to a HPLC system and mass spectrometer. In order to achieve high-throughput sample manipulation and analysis, a total of three HPLC pumps, two micro-digestion columns, and two C_8 separating columns were required. This allowed for continuous operation of the mass spectrometric system for analysis of the digested protein samples. Using the system, successful digests of cytochrome c and myoglobin were performed, and analysed by ESI-MS/MS in an automated fashion.

In Chapter 5, the automated online sample manipulation system was applied to a complex mixture consisting of *E. coli* proteins. Proteins were successfully extracted from *E. coli* and the whole cell lysate was fractionated by anion exchange chromatography. Portions of these fractions were subsequently loaded onto microcolumns for digestion in a continuous, fully automated fashion. A total of 68 proteins from the *E. coli* cell lysate were identified with the system. Thus, a sample manipulation system was created with high throughput capabilities.

It was determined in Chapter 5 that greater separation was required in order to identify more proteins. Two-dimensional HPLC has the potential resolving power required prior to digestion. One major problem with ion exchange relates to the poor resolution, resulting in proteins eluting over several (fractions) minutes. This was evident for *E. coli* proteins as seen in Chapter 5. If fractions are collected from a gradient elution, individual proteins are divided between several fractions. However, if proteins are eluted from the column *via* a step gradient, minimal amount of proteins will be divided among the fractions. Although the number of proteins per fraction would increase, the second mode of separation, namely reversed-phase chromatography, could separate these proteins. Each step in the salt gradient could be collected and injected into a reversed-phase column. If 10 gradient steps were employed in the separation, then a total of 10 samples would require reversed-phase separation. This second dimension of separation could theoretically produce approximately 50 fractions per sample. Thus, a total of 500 fractions would be generated, each requiring digestion and analysis. Sample manipulation of 500 samples by traditional methods would require many laborious hours, as well as high cost associated with, for example, 500 ZipTips for concentration. Using the technology described in this thesis, the total analysis time for these fractions would be 28 days. A flow chart of this strategy is presented in Figure 6.1. The time required for the mass spectrometry analysis of 500 samples using one C_8 column is approximately 54 days. The column must be washed and equilibrated prior to each run, resulting in doubling the analysis time. Still under the traditional method, if two columns were used in an automated set-up the mass spectrometry time would be 28 days; however, the time for sample preparation and the use of ZipTips are undesirable. The obvious advantage of this strategy is the relatively simple samples requiring digestion. Following two dimensions of separation, the number of proteins in each fraction would be significantly decreased. These fractions would be diluted after each chromatography stage. However, the system described here concentrates the samples prior to digestion.

A second strategy for increasing the number of proteins identified in a proteome involves switching the order of separations from the method described above. First, the whole cell lysate could be separated by reversed-phase chromatography. Gradient elution would produce fractions containing organic solvent. The organic portion would be removed prior to ion exchange separation in the second dimension. Coupling the ion exchange column to my sample manipulation system would enable further automation. However the flow rate from an ion exchange column would exceed the rate required for adsorption onto the reversed-phase microcolumn. Thus, a reservoir is required to store the effluent, then an HPLC pump using lower flow rates would be used to load the stored fraction onto the microcolumn for subsequent digestion. This could be achieved by

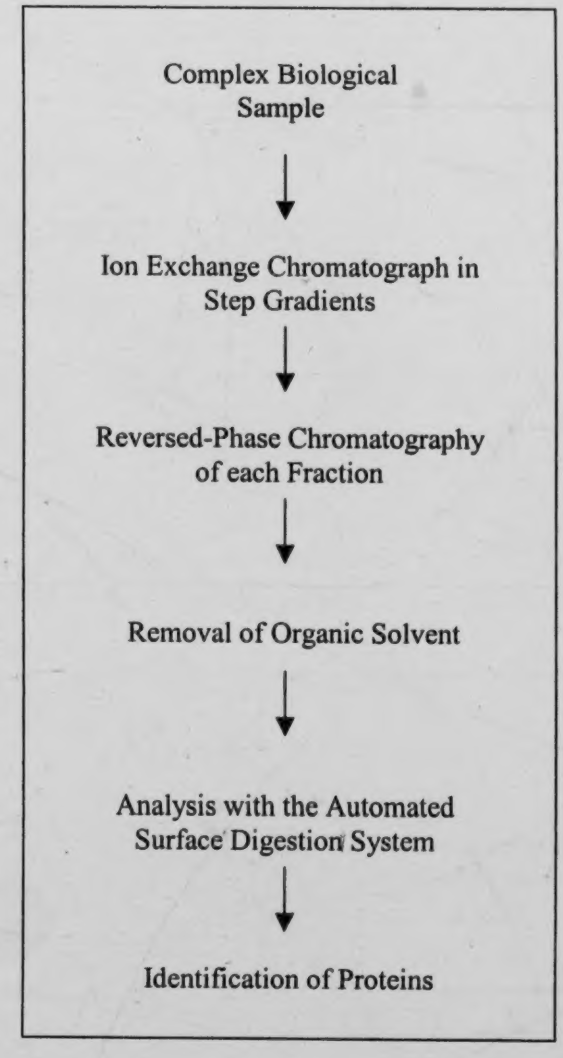

Complex Biological Sample
Ion Exchange Chromatograph in Step Gradients
Reversed-Phase Chromatography of each Fraction
Removal of Organic Solvent
Analysis with the Automated Surface Digestion System
Identification of Proteins

Figure 6.1 Flow chart representing a strategy for protein identification of complex samples.

adding PEEK tubing with a large i.d. to serve as the reservoir, two 6-port switching valves, and one HPLC pump to the previous system. A proposed automation for the ion exchange separation coupled with the sample manipulation system is presented in Figure 6.2. The effluent from the ion exchange column could be stored in [illegible] reservoir. HPLC pump 1, from the previous system, could then load the sample from the reservoir onto a microcolumn for subsequent digestion. Preceding the loading stage the sample manipulation is carried out in the same way as described in Chapter 4. In this set-up the proteins would be eluted from the ion exchange column in step gradients. The advantage of this method is that once the reversed-phase separation is completed and the organic solvent is removed from the fractions, the entire process is automated. Preliminary experiments have been performed using the reservoir for storing sample. Cytochrome c and chloropheniramine were stored in a 1 mL reservoir. The analytes were then loaded onto a reversed-phase microcolumn for concentration. The effluents of these analytes from the column were detected with a UV detector at 214 nm. The analytes were successfully concentrated onto the reversed-phase column. Therefore, the reservoir is capable of storing samples prior to loading onto a microcolumn. This strategy would reduce the number of proteins in each fraction, allowing for increased protein identification.

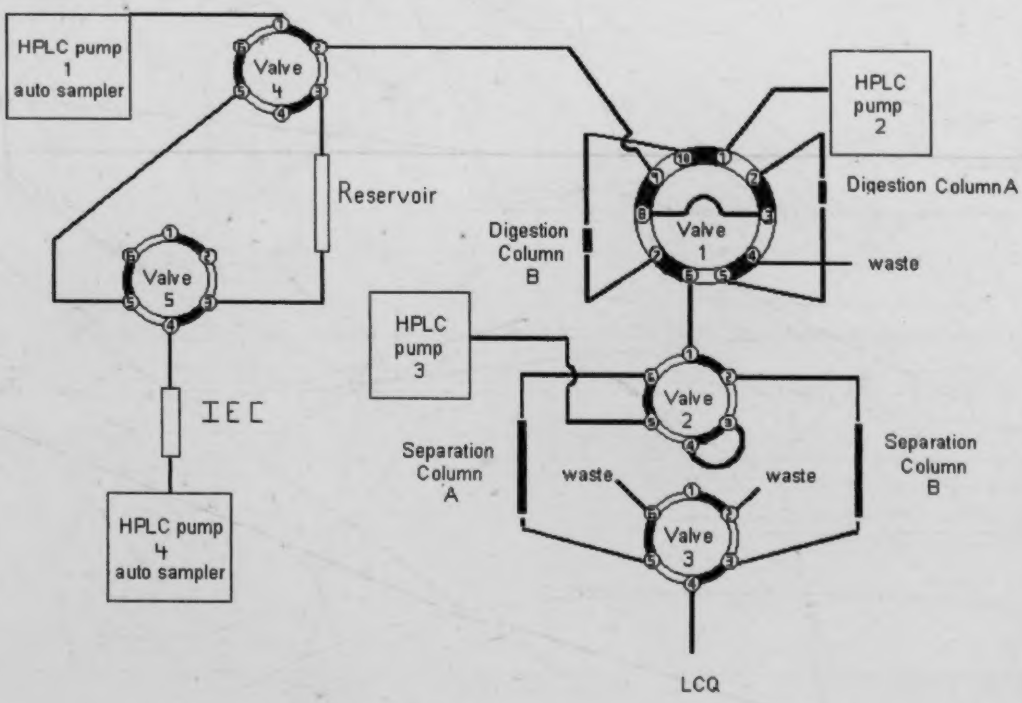

HPLC pump
1
auto sampler
Valve
4
Reservoir
HPLC pump
2
Digestion Column A
Digestion Column B
Valve
1
waste
Valve
5
HPLC pump
3
IEC
Valve
2
Separation Column A
Separation Column B
waste
waste
Valve
3
HPLC pump
4
auto sampler
LCQ

Figure 6.2 With the addition of two switching valves and one pump the ion exchange column can be directly coupled to the automated digestion system.

Regardless of the type of strategy used for proteome characterization, all systems must initially rely on a successful protein extraction step prior to analysis. However, given the different characteristics of proteins present in a cell, not all proteins can be extracted in one single step. Multiple extractions under different conditions, such as using extraction buffers over a wide pH range, should therefore be used to analyse the complete proteome. Also surfactants would be employed to solubilize and extract non-water soluble proteins, such as hydrophobic membrane proteins. There are many different techniques used to extract subcellular fractions of proteins. These must be explored if all proteins are going to be identified.

The automated system for protein digestion and separation was applied to a whole cell extract. Another direction of protein identification involves direct MALDI analysis of tissue samples. Over the past several years, investigators have demonstrated that proteins can be effectively desorbed directly from cells and tissues using MALDI-MS [2]. A review of single-cell MALDI published in 2000 provides an overview of the techniques and summarizes much of the work that has been done primarily on cells using MALDI-MS [3]. Direct analyses from tissues include the detection of proteins from rat pituitary, rat pancreas, mouse brain, human brain tumor xenografts, and mouse prostate. In addition, small molecular weight pharmaceutical compounds have also been directly detected in tissue. The direct proteomic analysis of tissue sections by MALDI-MS using microbeads for protein capturing is an emerging technique of great potential.

To capture proteins from a tissue, various different types of beads will be explored. In all studies presented in this thesis, proteins were captured with reversed-phase beads. However, many other types of beads could be used to capture proteins. In previous experiments performed in our lab, immobilized metal ion affinity chromatography (IMAC) beads were successfully used to concentrate and digest proteins. Other beads, such as anion, cation, normal-phase, and polymeric beads with specific chemical modifications could be explored. These beads could be used to selectively capture and concentrate specific proteins. Therefore, future efforts will concentrate on using different beads to selectively concentrate proteins in solutions.

MALDI-TOF-MS is well suited for the analysis of tissue samples with a spatial resolution of about 30 μm for compounds in a mass range from 1000 to over 50,000 Da. Currently the analysis of tissue focuses on detecting intact proteins. However, using my technology, proteins from tissue can be concentrated onto hydrophobic micro beads. These beads can then be removed from the tissue, where the adsorbed proteins can be digested and detected by mass spectrometry. In a preliminary experiment, HeLa cells, an adherent human cell line, were grown to confluence in a petri dish (compliments of Alan Doucette in the Carol Cass lab, Department of Oncology, University of Alberta). Poros R2 beads were placed on the cells in pure water. Water was used in order to swell the cells causing them to rupture. The resulting proteins were allowed to adsorb for 10 minutes. The beads were then removed from the petri dish and placed directly onto a MALDI target. Figure 6.3 displays the MALDI spectrum acquired from the beads. In a separate experiment the proteins adsorbed to the

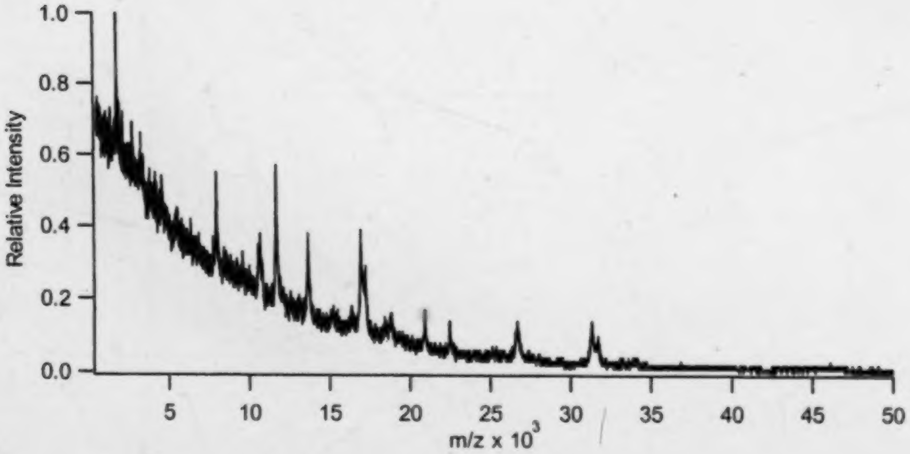


Figure 6.3 Direct MALDI spectrum of proteins adsorbed to Poros R2 beads from human cells.

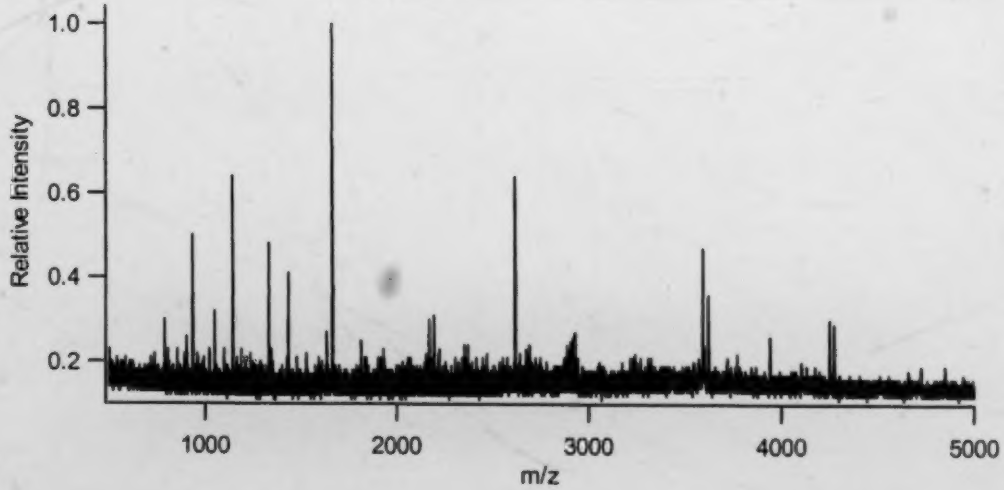


Figure 6.4 Direct MALDI spectrum of peptides produced from digesting proteins extracted from human cells on Poros R2 beads.

beads were digested with trypsin. Figure 6.4 displays the MALDI spectrum of the digest. There are several peptides detected from the digest. This data is encouraging, as it suggests that further analysis could identify several proteins from the generated peptides. However, a current limitation of this method is that it does not match the resolving power afforded by direct MALDI, which has a spatial resolution of 30 μm. To compete with this spatial resolution, continuing efforts are focused towards single bead analysis for tissue samples. This would significantly increase the resolution to the size of the beads being used.

In summary, the methods and system described in this thesis provide a means of not only handling high throughput analysis but also providing reliable protein identification. This is valuable technology, seeing as many current research efforts are focused towards proteome analysis. Although there are still a few shortcomings, future work, some of which was described in this chapter, can potentially improve upon this technology.

Literature Cited

1. Doucette, A.; Craft, D. R.; Li, L. *Anal. Chem.* **2000,** *72*, 3355-3362.
2. Chaurand, P.; Caprioli, R. M. *Electrophoresis* **2002,** *23*, 3125-3135.
3. Li, L.; Garden, R. W.; Sweedler, J. V. *Trends Biotechnol.* **2000,** *18*, 151-160.

Appendix A

DelCn - Delta Correlation value indicates how different the first hit is from the second hit in your search results. DelCn of 0.1 or greater is good.

Sp - Preliminary Score. The scoring is based on the number of ions in the MS/MS spectrum that match with the experimental data. The higher the value of Sp the better.

RSp - Ranking of the particular match during the preliminary scoring (see Sp above). Ideally, you'd like to see your match ranked #1 in the preliminary scoring.

Ions - The Ions value indicates how many of the experimental ions matched with the theoretical ions for the peptide listed. For example, 8/10 says that the MS/MS spectrum contained 8 of the 10 predicted ions for the peptide. It is rare to see 100% coverage of the predicted ions, but 70% or 80% coverage is good.

Xcorr - The Xcorr value is the cross-correlation value from the search. The #1 hit will always have the highest value of Xcorr, as Xcorr is used to produce the final ranking of the candidate peptides in the search. Usually, you'll see the top 10 or 12 ranked in your search results. XCorr values above 2.0 are usually indicative of a good correlation.